计算数学丛书

广义逆矩阵的基本理论和计算方法

何旭初

上海科学技术出版社

科技新书目：107·204
统一书号：13119·1277
定　　价：1.25 元

计算数学丛书

广义逆矩阵的基本理论和计算方法

何旭初

上海科学技术出版社

内 容 提 要

本书的目的是向读者介绍广义逆矩阵的基本理论以及计算广义逆矩阵和线性最小二乘问题的数值方法。全书共分五章，第1章的内容是有关线性代数的预备知识；第2章介绍了广义逆矩阵的基本理论；第3章着重讨论广义逆矩阵的连续性问题以及消除亏秩矩阵广义逆的不连续性的理论和方法；第4章则对奇异性和病态性这两个重要概念以及对度量病态程度的有效方法作了比较系统的介绍；第5章介绍了计算广义逆矩阵和线性最小二乘问题的数值方法，如LU分解，QR、QU分解以及直交化方法和递推方法等，章末还对有特殊结构的大型稀疏线性最小二乘问题的计算方法作了简要介绍。此外，在附录里对Drazin逆矩阵的基本理论也作了简单的介绍。

本书可供大学理、工科高年级学生、研究生、教师以及具有初步线性代数基础的工程技术人员参考。

计算数学丛书

广义逆矩阵的基本理论和计算方法

何旭初

上海科学技术出版社出版

（上海瑞金二路450号）

新华书店上海发行所发行　浙江湖州印刷厂印刷

开本787×1092 1/32　印张6.75　字数147,000

1985年11月第1版　1985年11月第1次印刷

印数：1—5,600

统一书号：13119·1277　　定价：1.25元

出 版 说 明

《计算数学丛书》是为了适应计算数学和计算机科学的发展，配合高等院校的计算数学教学的需要而组织的一套参考读物。读者对象主要是高等院校数学系和计算机科学系的学生、研究生，亦可供高等院校数学系和计算机科学系的教师以及工矿企业、科研单位从事计算工作的技术人员参考。

本丛书向读者介绍近代计算方法的一些主要进展及其适用范围和实用效果。每种书集中介绍一个专题，针对本专题的近代发展作综合性的介绍，内容简明扼要，重点突出，有分析，有评价，力图使读者对该专题的动向和发展趋势得到一个完整的了解。

本丛书已拟定的选题计有:《线性代数与多项式的快速算法》、《数论变换》、《数值有理逼近》、《矩阵特征值问题》、《索伯列夫空间引论》、《计算组合数学》、《样条与插值》、《不动点算法》、《广义逆矩阵的基本理论和计算方法》、《非线性方程的区间算法》、《奇异摄动中的边界层校正法》、《沃尔什函数理论与应用》、《多项式最佳逼近的实现》、《曲线曲面的数值表示和逼近》、《舍入误差分析引论》、《解边值问题的伽辽金法》、《非线性方程组迭代解法》、《外推法及其应用》、《蒙特卡罗方法》、《演化方程的有限元理论》、《数值解高维偏微分方程的分裂法》等二十余种，于一九八〇年初起陆续出版。

前　言

早在二十年代初期，E. H. Moore 就提出了广义逆矩阵这个重要概念．在三十年代，曾远荣先生把它推广到 Hilbert 空间中的线性算子上去，以后他又作了一系列有意义的工作．但是，长期以来，对广义逆矩阵的研究却没有受到人们的注意．随着科学技术的发展，直到五十年代中期，R. Penrose 又独立地提出广义逆矩阵的概念后，情况才开始发生了变化，这被人们称之为广义逆矩阵的再生．这是历史的必然．由于广义逆矩阵在测量学、统计学、经济学以及数学规划等许多领域中的重要应用逐渐为人们所认识，从而对广义逆矩阵的理论与应用的研究，产生了巨大的推动力量，使得这一学科在近三十年来得到迅速的发展．现在，广义逆矩阵的理论已成为线性代数中的一个重要方面，它是处理线性数学模型的一种有力工具．在计算数学中，特别由于它具有处理奇异性问题的能力，也日益显示出不可忽视的作用．

本书的目的是向读者介绍广义逆矩阵的基本理论以及计算广义逆矩阵和有关线性最小二乘问题的数值方法．全书共分五章，各章要点如下：

第 1 章是预备知识，主要介绍有限维向量空间及定义在其上的线性算子的基本理论．空间的直接和分解及由此确定的投影算子为讨论广义逆算子所必需．最后介绍了线性算子的矩阵表示，从而阐明了线性算子和矩阵之间的关系．

第 2 章介绍了广义逆矩阵的算子理论，以空间分解为基

础，给出了广义逆算子的构造性定义并讨论了它们的特征性质．通过广义逆算子的矩阵表示，建立了广义逆矩阵的相应理论．

第3章的内容是广义逆矩阵的摄动理论，讨论了摄动对广义逆矩阵的影响，其目的是研究广义逆矩阵的连续性问题．本章证明了亏秩矩阵广义逆的不连续性可以用人为的方法加以消除，从而对广义逆矩阵的计算提供了必要的理论基础，最后还讨论了摄动对最小二乘问题的解的影响．

第4章的目的之一是阐明问题的奇异性和病态性这两个重要概念及其区别．另一目的是建立判断问题病态程度的理论和方法．为此，本章对向量系的线性相关性理论作了定量研究，在此基础上讨论了问题的病态程度和相关性指标的关系，并提出了估计病态程度的拟条件数概念以及切实可行的计算方法．

第5章的内容是计算广义逆矩阵和相应线性最小二乘问题的计算方法．其中包括LU分解，QR和QU分解，$QR-LL^T$分解，以及直交化方法和递推方法等．最后还介绍了解大型稀疏线性最小二乘问题的计算方法．

鉴于Drazin逆矩阵在奇异线性常微分方程组以及奇异线性差分方程组等方面有重要应用，在附录里介绍了Drazin逆的基本理论．

罗亮生同志帮助绘制全书附图，赵金熙同志代为誊写，特别是孙麟平同志阅读了全稿并提出了一些有益的建议，作者谨向他们表示衷心的感谢！

何旭初　1983年5月于南京大学

符号说明

$\mathscr{V}, \mathscr{W}, \mathscr{U}, \cdots$ 表示向量空间

$\mathscr{C}^n, \mathscr{C}^m$ 表示 n 维和 m 维酉空间

$\mathscr{R}^n, \mathscr{R}^m$ 表示 n 维和 m 维欧氏空间

$\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{P}, \boldsymbol{Q}, \cdots$ 表示线性算子

$A, B, P, Q, \cdots$ 表示矩阵

$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}, \cdots$ 表示向量

$\mathscr{N}(\boldsymbol{A})$表示线性算子 $\boldsymbol{A}$ 的零空间

$\mathscr{R}(\boldsymbol{A})$表示线性算子 $\boldsymbol{A}$ 的像空间

$\mathscr{N}^{\perp}(\boldsymbol{A}), \mathscr{R}^{\perp}(\boldsymbol{A})$表示相应空间的直交补空间

$\dim \mathscr{V}$ 表示向量空间的维数

$\mathrm{rank}(\boldsymbol{A}), \mathrm{rank}(A)$表示相应算子和矩阵的秩

$k(A)$表示 A 的谱条件数

目　录

前言

符号说明

第 1 章　线性空间和线性算子 …………………………………1

§ 1　线性向量空间 …………………………………………1

§ 2　子空间和空间的分解 …………………………………8

§ 3　欧氏空间和酉空间 ……………………………………12

§ 4　线性算子及其矩阵表示 ………………………………18

§ 5　酉空间上的线性算子 …………………………………26

§ 6　投影算子及其矩阵表示 ………………………………32

§ 7　向量和算子的范数 ……………………………………41

第 2 章　广义逆矩阵的算子理论 …………………………47

§ 1　线性算子方程的求解问题 ……………………………47

§ 2　相容方程求解问题和相应的广义逆算子 A^- ……………48

§ 3　相容方程的极小范数解和广义逆算子 A_m^- ……………53

§ 4　矛盾方程的最小二乘解和广义逆算子 A_l^- ……………56

§ 5　矛盾方程的极小最小二乘解和广义逆算子 A^+ …………62

§ 6　广义逆算子的矩阵表示 ………………………………70

§ 7　一些特殊矩阵的 Moore-Penrose 广义逆 ………………77

第 3 章　广义逆矩阵的摄动理论和连续性问题 ……………80

§ 1　矩阵的奇异值分解和奇异值摄动定理 ………………80

§ 2　Banach 引理和非奇异方阵的摄动定理 ………………89

§ 3　满秩矩阵广义逆的连续性问题 ………………………92

§ 4　亏秩矩阵广义逆的连续性问题 ………………………95

§ 5　亏秩矩阵的保秩变形和广义逆矩阵不连续性的消除 ……101

§ 6　最小二乘问题的摄动定理 ……………………………………105
第 4 章　病态问题和病态程度的度量 ……………………109
§ 1　病态问题和算法的数值稳定性 ……………………………109
§ 2　数值相关性理论 ……………………………………………123
§ 3　计算相关性指标的方法 ……………………………………133
§ 4　矩阵的奇异度、条件数和伪秩……………………………138
第 5 章　广义逆矩阵和线性最小二乘问题的计算方法 …144
§ 1　计算广义逆矩阵的基本原则 ………………………………144
§ 2　Gauss 消去法 …………………………………………………145
§ 3　QR 或 QU 分解法……………………………………………152
§ 4　直交化方法 ……………………………………………………164
§ 5　计算广义逆矩阵和极小最小二乘解的递推方法 …………168
§ 6　大型稀疏线性最小二乘问题的计算方法 …………………175
附录 1　方阵的 Drazin 逆……………………………………………181
附录 2　关于广义逆乘积公式证明的补充 …………………………199
参考文献 ……………………………………………………………200
索引 …………………………………………………………………204

第 1 章

线性空间和线性算子

§1　线性向量空间

1-1　线性向量空间的定义

设 $\mathscr{V}$ 为某一集合，其元素 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \cdots$ 等叫做向量．在 $\mathscr{V}$ 的元素之间，定义有加法运算，记作"+"．这种加法运算具有下列性质：

1）$\mathscr{V}$ 中的任意两个向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 都可以相加，其和仍为 $\mathscr{V}$ 中的一个向量 $\boldsymbol{z}$，记作

$$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{z},$$

这就是说，集合 $\mathscr{V}$ 对加法运算是封闭的；

2）加法运算服从交换律和结合律，即对 $\mathscr{V}$ 中任意的向量 $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$，恒有

$$\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}, \quad \text{交换律},$$

$$(\boldsymbol{x}+\boldsymbol{y})+\boldsymbol{z}=\boldsymbol{x}+(\boldsymbol{y}+\boldsymbol{z}), \quad \text{结合律};$$

3）$\mathscr{V}$ 中有唯一的元素 $\mathbf{0}$，称为零元素或零向量，它具有性质：对任意的向量 $\boldsymbol{x}\in\mathscr{V}$（即 $\boldsymbol{x}$ 属于 $\mathscr{V}$），

$$\boldsymbol{x}+\mathbf{0}=\mathbf{0}+\boldsymbol{x}=\boldsymbol{x};$$

4）对任意的向量 $\boldsymbol{x}\in\mathscr{V}$，总存在一个向量 $\boldsymbol{y}$，它具有性质：

$$\boldsymbol{x}+\boldsymbol{y}=\mathbf{0},$$

我们称 $\boldsymbol{y}$ 为 $\boldsymbol{x}$ 的逆元素，记作 $\boldsymbol{y}=-\boldsymbol{x}$．

其次，我们设 K 是一个给定的数域，例如复数域或实数域（分别记作 $\mathscr{C}$ 或 $\mathscr{R}$），其中的数记作 $\alpha, \beta, \gamma, \cdots$．$K$ 中的

数 α 和 $\mathscr{V}$ 中的向量 $\boldsymbol{x}$ 之间定义有乘法，记作 $\alpha\boldsymbol{x}$，它也是 $\mathscr{V}$ 中的一个向量. 这种运算满足下列规律:

对任意的数 $\alpha, \beta \in K$，和任意的向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{V}$，恒有:

5) $\alpha(\beta\boldsymbol{x})=(\alpha\beta)\boldsymbol{x}$，　结合律;

6) $(\alpha+\beta)\boldsymbol{x}=\alpha\boldsymbol{x}+\beta\boldsymbol{x}$，　对于数的分配律;

7) $\alpha(\boldsymbol{x}+\boldsymbol{y})=\alpha\boldsymbol{x}+\alpha\boldsymbol{y}$，　对于向量的分配律;

8) $1\boldsymbol{x}=\boldsymbol{x}$.

定义 1　具有上述性质 1)～8)的集合 $\mathscr{V}$，称为域 K 上的**线性向量空间**，或简称之为**向量空间**.

例 1　关于变元 t 的一切不高于 $n-1$ 次的实系数(或复系数)多项式

$$p_n(t)=a_0t^{n-1}+a_1t^{n-2}+\cdots+a_{n-1}$$

所组成的集合，按通常的意义进行多项式和数的乘法以及多项式的加法运算，则此集合便是一个实(或复)数域上的向量空间，记作 $\mathscr{P}_n$.

例 2　一切 n 元有序的实(或复)数组

$$(x_1,\ x_2,\ \cdots,\ x_n)$$

所组成的集合 $\mathscr{R}^n$，若规定相应的"加法"和"乘法"运算为:

$$(x_1,\ \cdots,\ x_n)+(y_1,\ \cdots,\ y_n)=(x_1+y_1,\ \cdots,\ x_n+y_n),$$

$$\alpha(x_1,\ \cdots,\ x_n)=(\alpha x_1,\ \cdots,\ \alpha x_n),$$

这种集合便构成实(或复)数域上的向量空间，常记作 $\mathscr{R}^n$(或 $\mathscr{C}^n$).

例 3　一切收敛的实(或复)数序列 $\{x_i\}$，按通常的意义规定"加法"和"乘法"运算，所组成的集合也是相应数域上的向量空间. 对于这种情形，应补充规定: 若 $\lim x_i=\lim y_i$，则规定二序列 $\{x_i\}$ 和 $\{y_i\}$ 为相等.

1-2 向量系的线性相关性

设 $\mathscr{V}$ 是定义在数域 K 上的向量空间，$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 为其中某 n 个元素. 我们有

定义 2 若存在不全为零的数 $\alpha_i \in K$, $i=1, \cdots, n$, 使等式

$$\alpha_1\boldsymbol{x}_1+\cdots+\alpha_n\boldsymbol{x}_n=\mathbf{0}$$

成立，便称向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 为**线性相关**，否则的话，便称之为线性独立或**线性无关**.

应当指出，对任意的向量 $\boldsymbol{x}\in\mathscr{V}$，恒有

$$0\boldsymbol{x}=\mathbf{0},$$

因为我们总有

$$1\boldsymbol{x}+0\boldsymbol{x}=(1+0)\boldsymbol{x}=\boldsymbol{x}.$$

今将有关线性相关向量系的一些简单事实列举于下：

1° 若向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 中有一个为零向量，则此向量系必为线性相关，反之则不然。

2° 若向量系中有一个向量，例如设其为 $\boldsymbol{x}_n$，可以表示成其余向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{n-1}$ 的线性组合，即

$$\boldsymbol{x}_n=\alpha_1\boldsymbol{x}_1+\cdots+\alpha_{n-1}\boldsymbol{x}_{n-1},$$

则向量系必为线性相关，因为显然有

$$\alpha_1\boldsymbol{x}_1+\cdots+\alpha_{n-1}\boldsymbol{x}_{n-1}-\alpha_n\boldsymbol{x}_n=\mathbf{0},$$

其中 $\alpha_n=-1\neq 0$.

3° 若 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 为线性相关，则其中至少有一个向量可以表示成其余向量的线性组合.

4° 若向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 的某一个子集，例如，可设其为 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r\}$, $r<n$, 为线性相关，则原向量系也必为线性相关.

1-3 向量空间的基底和维数

为了给出向量空间的表示，我们先引进空间基底这个重要概念.

定义3 设 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 为向量空间(K 域上的) $\mathscr{V}$ 中的一个线性独立系，如果空间中的任何向量 $\boldsymbol{x}$ 都可以表示成 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 的线性组合，即，存在数 $\alpha_i \in K$, $i=1, \cdots, n$, 使

$$\boldsymbol{x} = \alpha_1\boldsymbol{x}_1 + \cdots + \alpha_n\boldsymbol{x}_n, \tag{1}$$

则称向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 构成空间 $\mathscr{V}$ 的一个基底.

定理1 设 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 构成空间 $\mathscr{V}$ 的一个基底，则表达式(1)是唯一的.

证明 设 $\boldsymbol{x}$ 有另一个表达式

$$\boldsymbol{x} = \beta_1\boldsymbol{x}_1 + \cdots + \beta_n\boldsymbol{x}_n, \tag{2}$$

(1)与(2)相减得

$$(\alpha_1 - \beta_1)\boldsymbol{x}_1 + \cdots + (\alpha_n - \beta_n)\boldsymbol{x}_n = \boldsymbol{0},$$

因为向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 为线性独立，所以必有

$$\alpha_i - \beta_i = 0, \quad i=1, \cdots, n,$$

即 $\alpha_i = \beta_i$, $i=1, \cdots, n$, 这就证明了表达式(1)的唯一性. 证毕.

进一步我们还有

定理2 设向量系 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 和 $\{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_m\}$ 都是向量空间 $\mathscr{V}$ 的基底，其中 n 和 m 是两个自然数，则必有 $m=n$.

证明 假定 $m>n$. 因为 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 是 $\mathscr{V}$ 的一个基底，所以每一个 $\boldsymbol{y}_i$, $i=1, \cdots, m$ 都可以表示成它们的线性组合:

$$\begin{aligned} \boldsymbol{y}_1 &= a_{11}\boldsymbol{x}_1 + a_{21}\boldsymbol{x}_2 + \cdots + a_{n1}\boldsymbol{x}_n, \\ \boldsymbol{y}_2 &= a_{12}\boldsymbol{x}_1 + a_{22}\boldsymbol{x}_2 + \cdots + a_{n2}\boldsymbol{x}_n, \\ &\cdots\cdots\cdots\cdots \\ \boldsymbol{y}_m &= a_{1m}\boldsymbol{x}_1 + a_{2m}\boldsymbol{x}_2 + \cdots + a_{nm}\boldsymbol{x}_n. \end{aligned} \tag{3}$$

今证明必存在不全为零的数 $\alpha_1, \cdots, \alpha_m$ 使

$$\alpha_1\boldsymbol{y}_1 + \alpha_2\boldsymbol{y}_2 + \cdots + \alpha_m\boldsymbol{y}_m = \boldsymbol{0}. \tag{4}$$

为此，我们把 $\boldsymbol{y}_i$ 的表达式(3)代入(4)便得到:

$$(a_{11}\alpha_1+a_{12}\alpha_2+\cdots+a_{1m}\alpha_m)\boldsymbol{x}_1$$
$$+(a_{21}\alpha_1+a_{22}\alpha_2+\cdots+a_{2m}\alpha_m)\boldsymbol{x}_2$$
$$+\cdots+(a_{n1}\alpha_1+a_{n2}\alpha_2+\cdots+a_{nm}\alpha_m)\boldsymbol{x}_n=\mathbf{0}.$$

令 $\boldsymbol{x}_i(i=1,\cdots,n)$的系数为零，便得到齐次线性方程组

$$a_{11}\alpha_1+a_{12}\alpha_2+\cdots+a_{1m}\alpha_m=0,$$
$$a_{21}\alpha_1+a_{22}\alpha_2+\cdots+a_{2m}\alpha_m=0,$$
$$\cdots\cdots\cdots\cdots$$
$$a_{n1}\alpha_1+a_{n2}\alpha_2+\cdots+a_{nm}\alpha_m=0.$$

这是一个亚定方程组，即方程个数少于未知量 α_1、$\cdots$、α_m 的个数，所以必有非零解．这就是说，存在有不全为零的数 $\alpha_1,\cdots,\alpha_m$ 使(4)成立，从而可知，$\boldsymbol{y}_1,\cdots,\boldsymbol{y}_m$ 必为线性相关，因此必有

$$m\leqslant n.$$

交换 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_j$，可证 $n\leqslant m$，于是必有

$$m=n.$$

证毕.

向量空间的基底不是唯一的，但是，任何基底中的向量个数则必相等．于是我们有

定义 4 向量空间 $\mathscr{V}$ 若存在有限基底 $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$，则规定空间 $\mathscr{V}$ 的维数为 n，记作

$$\dim\mathscr{V}=n.$$

若不存在有限基底，则称相应空间的维数为无穷．

例如，在多项式空间 $\mathscr{P}_n$ 中，$1, t, \cdots, t^{n-1}$ 便构成一个基底，所以

$$\dim\mathscr{P}_n=n.$$

又如，$(1,0,\cdots,0)$，$(0,1,0,\cdots,0)$，$\cdots$，$(0,\cdots,1)$为 n 个 n 元数组，它们构成 $\mathscr{R}^n$(或 $\mathscr{C}^n$)的一个基底，所以

$$\dim\mathscr{R}^n=n.$$

又例如，前面所讲的收敛数列空间则是无穷维的．

本书只考虑有限维的情形．

1-4 向量空间的表示

我们首先考虑空间 $\mathscr{V}$ 的基底存在性问题. 设 $\boldsymbol{x}_1$ 是 $\mathscr{V}$ 中的一个非零向量,若 $\mathscr{V}$ 中的任何向量 $\boldsymbol{x}$ 都可以表示成

$$\boldsymbol{x}=\alpha\boldsymbol{x}_1$$

的形式,则 $\boldsymbol{x}_1$ 便是 $\mathscr{V}$ 的一个基底,这时

$$\dim\mathscr{V}=1.$$

不然的话,即存在非零向量 $\boldsymbol{x}_2$,它不能表示成上述形式. 但是,如果 $\mathscr{V}$ 中的任何向量 $\boldsymbol{x}$ 都可以表示成 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$ 的线性组合,即

$$\boldsymbol{x}=\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2,$$

则 $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ 便构成 $\mathscr{V}$ 的一个基底,这时空间的维数为 2. 否则的话,必存在非零向量 $\boldsymbol{x}_3$,它不能表示成上述形式,如此等等. 最后必有下列两种情形之一发生:

1) 存在有限个向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$,它们构成 $\mathscr{V}$ 的一个基底,这时 $\dim\mathscr{V}=n$;

2) 不存在有限基底,这时空间 $\mathscr{V}$ 是无限维的.

现在我们假定

$$\dim\mathscr{V}=n,$$

n 是一个自然数,这时存在基底

$$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n.$$

设 $\boldsymbol{x}$ 为 $\mathscr{V}$ 中的任一向量,则它必可唯一地表示成 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 的线性组合,即存在唯一的一组数 $\xi_1, \xi_2, \cdots, \xi_n$,使

$$\boldsymbol{x}=\xi_1\boldsymbol{x}_1+\xi_2\boldsymbol{x}_2+\cdots+\xi_n\boldsymbol{x}_n. \tag{5}$$

这时我们便称 $(\xi_1, \xi_2, \cdots, \xi_n)$ 为向量 $\boldsymbol{x}$ 在基底 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 下的表示,并称 ξ_i, $i=1, 2, \cdots, n$ 为向量 $\boldsymbol{x}$ 的第 i 个分量或第 i 个坐标.

这就是说,在引进了空间基底 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 以后,我们便

可以把空间中的向量表示成 n 元数组. 这也就是说，我们用 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 作为基底，在向量空间 $\mathscr{V}$ 中建立了一个坐标系. 这样，我们便把 n 维向量空间 $\mathscr{V}$ 用一个 n 元数组空间 $\mathscr{R}^n$(或 $\mathscr{C}^n$)表示出来.

应当注意，同一个向量，对于空间中不同的基底，它的坐标也不一样.

在基底 $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ 下，我们也称$(\xi_1, \cdots, \xi_n)$为 $\boldsymbol{x}$ 的坐标向量，也可以称 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 为这个坐标系的坐标轴. 向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 自身，在这个坐标系中的表示为:

$$\boldsymbol{x}_1=(1, 0, \cdots, 0), \quad \boldsymbol{x}_2=(0, 1, 0, \cdots, 0),$$
$$\cdots\cdots\cdots\cdots \quad \boldsymbol{x}_n=(0, 0, \cdots, 1).$$

所以也可以称右端为空间 $\mathscr{V}$ 的单位坐标向量.

1-5　空间的同构

根据上面的讨论可以看出，n 维向量空间 $\mathscr{V}$，在建立了坐标系后便和同一数域 K 上的 n 元数组空间 K^n 之间建立了一种对应关系. 这种对应关系不但是一对一的，而且还保持关于向量运算的线性性质，即若 $\boldsymbol{x}, \boldsymbol{y}\in\mathscr{V}$ 分别和 K 域上的 n 元数组$(\xi_1, \xi_2, \cdots, \xi_n)$，$(\eta_1, \eta_2, \cdots, \eta_n)$对应:

$$\boldsymbol{x}\leftrightarrow(\xi_1, \cdots, \xi_n),$$
$$\boldsymbol{y}\leftrightarrow(\eta_1, \cdots, \eta_n),$$

对于 $\alpha, \beta\in K$，恒有

$$\begin{aligned}\alpha\boldsymbol{x}+\beta\boldsymbol{y}&\leftrightarrow(\alpha\xi_1+\beta\eta_1, \cdots, \alpha\xi_n+\beta\eta_n)\\&=\alpha(\xi_1, \cdots, \xi_n)+\beta(\eta_1, \cdots, \eta_n).\end{aligned} \tag{6}$$

我们称这种关系为向量空间 $\mathscr{V}$ 和同一数域 K 上的 n 元数组空间 K^n 之间的同构关系. 在一般情形，我们有

定义 5　设 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 为同一数域 K 上的向量空间，若存在双向单值映射 T，即，对任何 $\boldsymbol{x}\in\mathscr{V}_1$，对应有唯一的 $\boldsymbol{y}=$

$T\boldsymbol{x}\in\mathscr{V}_2$,并且,若 $\boldsymbol{x}_1, \boldsymbol{x}_2\in\mathscr{V}_1, \boldsymbol{x}_1\neq\boldsymbol{x}_2$,则

$$\boldsymbol{y}_1=T\boldsymbol{x}_1\neq T\boldsymbol{x}_2=\boldsymbol{y}_2,$$

并且,这种映射还保持空间向量运算的线性性质,即,对任意的 $\alpha, \beta\in K$,恒有

$$T(\alpha\boldsymbol{x}_1+\beta\boldsymbol{x}_2)=\alpha T\boldsymbol{x}_1+\beta T\boldsymbol{x}_2,$$

则称向量空间 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 为同构.

很明显,这种同构关系具有传递性,就是说,若 $\mathscr{V}_1, \mathscr{V}_2, \mathscr{V}_3$ 为同一数域 K 上的向量空间,且已知 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 同构,$\mathscr{V}_2$ 和 $\mathscr{V}_3$ 同构,则 $\mathscr{V}_1$ 必和 $\mathscr{V}_3$ 同构. 而且还不难证明,若 $\mathscr{V}_1, \mathscr{V}_2$ 同构,且 $\dim\mathscr{V}_1=n$,则必有

$$\dim\mathscr{V}_2=n,$$

即,同构的向量空间有相同的维数.

§2 子空间和空间的分解

2-1 子空间

设 $\mathscr{V}$ 是数域 K 上的有限维向量空间,S 是 $\mathscr{V}$ 的一个非空子集,它对向量运算具有封闭性,即,若 $\boldsymbol{x}, \boldsymbol{y}\in S, \alpha, \beta\in K$,则

$$\alpha\boldsymbol{x}+\beta\boldsymbol{y}\in S.$$

这一类型的子集具有十分重要的性质.

定义 6 具有上述特征的子集 S,称为向量空间 $\mathscr{V}$ 的一个子空间.

例如,设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k\in\mathscr{V}$,它们的一切线性组合

$$\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2+\cdots+\alpha_k\boldsymbol{x}_k$$

所构成的一个集合 S,便是一个子空间. 如果向量 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ 为线性独立,则此子空间的维数为 k,

$$\dim S=k.$$

这就是向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ 所张成的子空间，常记作

$$\operatorname{Span}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\}.$$

只包含零向量的子空间，记作 $\mathscr{O}$，其维数定义为 0.

例　在 $\mathscr{P}_n$ 中一切次数不高于 $r-1(r\leqslant n)$ 的变元 t 的多项式的全体，构成一个 r 维子空间，$1, t, \cdots, t^{r-1}$ 就是这个子空间的一组基底，所以，又可以表示成

$$\operatorname{Span}\{1, t, \cdots, t^{r-1}\}.$$

在 $\mathscr{R}^n$ 中，前 k 个分量固定为 0 的一切 n 元数组

$$(0, \cdots, \xi_{k+1}, \cdots, \xi_n), \quad k\leqslant n$$

的全体，构成一个 $n-k$ 维的子空间．若记

$$\boldsymbol{e}_1=(1, 0, \cdots, 0), \quad \boldsymbol{e}_2=(0, 1, 0, \cdots, 0)$$
$$\cdots\cdots\cdots, \quad \boldsymbol{e}_n=(0, 0, \cdots, 0, 1),$$

则上述子空间可以表示为

$$\operatorname{Span}\{\boldsymbol{e}_{k+1}, \cdots, \boldsymbol{e}_n\}.$$

给定数 $\alpha_1, \alpha_2, \cdots, \alpha_n$，满足线性齐次方程

$$\alpha_1\xi_1+\cdots+\alpha_n\xi_n=0$$

的一切 $(\xi_1, \cdots, \xi_n)$，构成 $\mathscr{R}^n$ 的一个子空间，其维数为 $n-1$，常称这种子空间为 $\mathscr{R}^n$ 中的一个 $n-1$ 维超平面.

2-2　子空间的运算

设 S_1, S_2 为包含于向量空间 $\mathscr{V}$（域 K 上的）的两个子空间，我们用记号 $S_1\cap S_2$ 表示它们的交，即共同部分．这就是说，$\boldsymbol{x}\in S_1\cap S_2$ 的充分必要条件为既要 $\boldsymbol{x}\in S_1$，又要 $\boldsymbol{x}\in S_2$，记

$$S=S_1\cap S_2, \tag{7}$$

则 S 也必为 $\mathscr{V}$ 的一个子空间.

因为，若 $\boldsymbol{x}, \boldsymbol{y}\in S$，则必有

$$\boldsymbol{x}, \boldsymbol{y}\in S_1, \quad \boldsymbol{x}, \boldsymbol{y}\in S_2,$$

由于 S_1, S_2 都是子空间，所以，对任意的 $\alpha, \beta\in K$，

$$\alpha\boldsymbol{x}+\beta\boldsymbol{y}\in S_1,\quad \alpha\boldsymbol{x}+\beta\boldsymbol{y}\in S_2,$$

从而必有 $$\alpha\boldsymbol{x}+\beta\boldsymbol{y}\in S,$$

即，S 也是一个子空间.

同理，k 个子空间 $S_1, S_2, \cdots, S_k$ 的交

$$S=S_1\cap S_2\cap\cdots\cap S_k$$

仍然是一个子空间.

交的概念可以推广到无限多个子空间的情形.

给定向量空间 $\mathscr{V}$ 的任一子集 $\mathscr{M}$，包含它的一切子空间的交，便是包含 $\mathscr{M}$ 的最小子空间. 例如，包含 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ 的一切子空间的交，实际上就是

$$\mathrm{Span}\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k\},$$

它是包含 k 个给定向量 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ 的最小子空间.

2-3 空间的分解

对于给定的子空间 S_1、S_2，记 S 为包含它们的最小子空间，则 S 为一切形如

$$\boldsymbol{u}=\boldsymbol{x}+\boldsymbol{y},\quad \boldsymbol{x}\in S_1,\ \boldsymbol{y}\in S_2 \tag{8}$$

的向量 $\boldsymbol{u}$ 的全体所构成的子空间，因而可以表示成

$$S=S_1+S_2. \tag{9}$$

如果表达式(8)是唯一的，则 S 便称为 S_1 和 S_2 的直接和，并记成

$$S=S_1\oplus S_2. \tag{10}$$

若向量空间 $\mathscr{V}$ 表示成直接和

$$\mathscr{V}=\mathscr{M}\oplus\mathscr{N}, \tag{11}$$

其中 $\mathscr{M}$ 和 $\mathscr{N}$ 是 $\mathscr{V}$ 的两个子空间，则它们为互补，即，$\mathscr{M}$ 为 $\mathscr{N}$ 的补空间，而 $\mathscr{N}$ 为 $\mathscr{M}$ 的补空间. 我们称(10)和(11)分别为子空间 S 和空间 $\mathscr{V}$ 的直接和分解.

定理 8 直接和分解(10)中的子空间 S_1 和 S_2 恰有一个

共同元素，即零向量.

证明 因为 S_1 和 S_2 都是 S 的子空间，所以都包含零向量，如果它们还有另一个共同向量，设其为 $\boldsymbol{v}$，即

$$\boldsymbol{v}\in S_1\cap S_2,$$

设 $\boldsymbol{u}=\boldsymbol{x}+\boldsymbol{y}\in S$, $\boldsymbol{x}\in S_1$, $\boldsymbol{y}\in S_2$，则 $\boldsymbol{u}$ 还可以表示成

$$\boldsymbol{u}=(\boldsymbol{x}+\boldsymbol{v})+(\boldsymbol{y}-\boldsymbol{v}),$$

从而 $\boldsymbol{u}$ 的表达式便不是唯一的了，除非 $\boldsymbol{v}=\boldsymbol{0}$. 证毕.

直接和分解的例 今考虑 n 元数组空间 $\mathscr{R}^n$，

令
$$\mathscr{R}_k=\{(\xi_1,\ \cdots,\ \xi_k,\ 0,\ \cdots,\ 0)\},$$
$$S_k=\{(0,\ \cdots,\ 0,\ \xi_{k+1},\ \cdots,\ \xi_n)\},$$

则
$$\mathscr{R}^n=\mathscr{R}_k\oplus S_k.$$

直接和分解还可以推广到更多子空间的情形，例如，

$$\mathscr{R}^n=\mathrm{Span}\{\boldsymbol{e}_1\}\oplus\mathrm{Span}\{\boldsymbol{e}_2\}\oplus\cdots\oplus\mathrm{Span}\{\boldsymbol{e}_n\},$$

其中的 $\boldsymbol{e}_i$ 是以前所讲的单位坐标向量.

2-4 维数关系

现在我们来考虑直接和分解式(10)中各子空间维数之间的关系.

定理 4 设(10)为空间 S 的一种直接和分解，且

$$\dim S=m,\ \dim S_1=r_1,\ \dim S_2=r_2,$$

则必有
$$m=r_1+r_2,$$

即
$$\dim S=\dim S_1+\dim S_2. \tag{12}$$

证明 因为 $\dim S_1=r_1$，所以 S_1 中存在基底 $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{r_1}$，同理，$S_2$ 中存在基底 $\boldsymbol{y}_1,\cdots,\boldsymbol{y}_{r_2}$. 根据直接和的定义可知，$\{\boldsymbol{x}_1,\ \cdots,\ \boldsymbol{x}_{r_1},\ \boldsymbol{y}_1,\ \cdots,\ \boldsymbol{y}_{r_2}\}$ 便构成空间 S 的一个基底，所以(12)成立. 证毕.

§3 欧氏空间和酉空间

3-1 向量的内积

在向量分析中，向量的数量积是一个重要的概念. 现在把它推广到抽象的向量空间中来. 设 $\mathscr{V}$ 是复数域 $\mathscr{C}$ 上的向量空间，$\boldsymbol{x}$，$\boldsymbol{y}$ 为空间中任意的向量. 用记号$(\boldsymbol{x}, \boldsymbol{y})$表示 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的一个复值函数，它具有下列性质：

1) $(\boldsymbol{x}, \boldsymbol{y})=\overline{(\boldsymbol{y}, \boldsymbol{x})}$，记号"——"表示取共轭复数；

2) $(\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2, \boldsymbol{y})=\alpha_1(\boldsymbol{x}_1, \boldsymbol{y})+\alpha_2(\boldsymbol{x}_2, \boldsymbol{y})$，其中，

$$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}\in\mathscr{V},\ \alpha_1, \alpha_2\in\mathscr{C};$$

3) $(\boldsymbol{x}, \boldsymbol{x})\geqslant 0$，对任何 $\boldsymbol{x}\in\mathscr{V}$，并且仅当 $\boldsymbol{x}=\boldsymbol{0}$ 时等式成立.

首先可知，因为关系 1)成立，所以，$(\boldsymbol{x}, \boldsymbol{x})$为实数，故性质 3)有意义.

定义 7 我们称具有上述性质的函数$(\boldsymbol{x}, \boldsymbol{y})$为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的内积. 并称

$$\sqrt{(\boldsymbol{x}, \boldsymbol{x})}$$

为向量 $\boldsymbol{x}$ 的范数或长度，记作 $\|\boldsymbol{x}\|^2=(\boldsymbol{x}, \boldsymbol{x})$.

我们称定义有上述内积和向量范数的向量空间是一个酉空间(Unitary Space).

对于定义在实数域 $\mathscr{R}$ 上的向量空间，可仿前定义内积如下. $(\boldsymbol{x}, \boldsymbol{y})$为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的实值函数，它具有下列性质：

1) $(\boldsymbol{x}, \boldsymbol{y})=(\boldsymbol{y}, \boldsymbol{x})$， 对称性；

2) $(\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2, \boldsymbol{y})=\alpha_1(\boldsymbol{x}_1, \boldsymbol{y})+\alpha_2(\boldsymbol{x}_2, \boldsymbol{y})$， 线性性质；

3) $(\boldsymbol{x}, \boldsymbol{x})\geqslant 0$，只有 $\boldsymbol{x}=\boldsymbol{0}$ 时等号才成立，这是内积的非负性.

向量 $\boldsymbol{x}$ 的范数仍规定为

$$\|\boldsymbol{x}\|^2=(\boldsymbol{x}, \boldsymbol{x}).$$

定义有上述内积和向量范数的实向量空间，称之为欧氏空间(Euclidean Space).

今后我们常把 n 维酉空间记成 $\mathscr{C}^n$，n 维欧氏空间记成 $\mathscr{R}^n$.

3-2 直交性和 Gram-Schmidt 直交化方法

在空间中引进了内积，就可以进一步引进直交性这个重要概念了. 今设 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 为酉空间 $\mathscr{C}^n$ (或欧氏空间 $\mathscr{R}^n$)中的向量，我们有

定义 8 若$(\boldsymbol{x}, \boldsymbol{y})=0$，便称向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 为直交，并记作

$$\boldsymbol{x} \perp \boldsymbol{y}.$$

显然，直交性是相互的，因为，若$(\boldsymbol{x}, \boldsymbol{y})=0$，则

$$\overline{(\boldsymbol{y}, \boldsymbol{x})}=(\boldsymbol{x}, \boldsymbol{y})=0,$$

从而必有

$$(\boldsymbol{y}, \boldsymbol{x})=0.$$

容易看出，零向量和任何向量都直交.

定理 5 设 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ 是空间 $\mathscr{C}^n$ (或 $\mathscr{R}^n$)中的一个非零直交系，即，$\boldsymbol{x}_i \neq 0$，$(\boldsymbol{x}_i, \boldsymbol{x}_j)=0$，对 $i \neq j$，$i, j=1, 2, \cdots, k$，则它们必为线性独立.

这就是说，非零直交系必为线性独立系.

证明 假定存在有线性关系

$$\alpha_1 \boldsymbol{x}_1+\cdots+\alpha_k \boldsymbol{x}_k=0,$$

则对任何 j 必有

$$0=\left(\sum_{i=1}^k \alpha_i \boldsymbol{x}_i, \boldsymbol{x}_j\right)=\sum_{i=1}^k \alpha_i(\boldsymbol{x}_i, \boldsymbol{x}_j)=\alpha_j(\boldsymbol{x}_j, \boldsymbol{x}_j).$$

因为 $\boldsymbol{x}_j \neq \mathbf{0}$，所以$(\boldsymbol{x}_j, \boldsymbol{x}_j)>0$，从而必有

$$\alpha_j=0, \quad j=1, \cdots, k.$$

这就证明了向量系 $x_1, \cdots, x_k$ 是线性独立的. 证毕.

定义 9 设 $x_1, \cdots, x_k$ 是 $\mathscr{C}^n$ 中的一个直交系,若

$$\|x_i\|=1, \quad i=1, 2, \cdots, k,$$

则称向量系 $x_1, \cdots, x_k$ 为 $\mathscr{C}^n$ 中的一个**标准直交系**.

现在我们进一步考虑空间 $\mathscr{C}^n$ 中标准直交系的存在性,从而为在空间中建立标准直交基底打下基础.

定理 6 在空间 $\mathscr{C}^n$ 中必存在标准直交系 $x_1, \cdots, x_n$, 即, $(x_i, x_j)=\delta_{ij}$,

$$\delta_{ij}=\begin{cases}1, & i=j,\\ 0, & i\neq j,\end{cases} \quad i, j=1, \cdots, n. \tag{13}$$

我们称 δ_{ij} 为 Kronecker delta.

证明 我们用一种构造性方法来证明标准直交系的存在性.

因为 $\dim \mathscr{C}^n=n$, 所以必存在基底

$$y_1, y_2, \cdots, y_n.$$

现在我们用这些 y_i 的线性组合来建立标准直交系 $\{x_i\}$. 首先令

$$x_1=y_1/\|y_1\|.$$

于是 $\|x_1\|=1$.

现在我们假定已构造标准直交系

$$x_1, x_2, \cdots, x_r,$$

其中 x_j 为 $y_1, \cdots, y_j$ 的线性组合, $j=1, 2, \cdots, r$. 今确定 x_{r+1} 如下:

令

$$u=y_{r+1}-(\alpha_1 x_1+\cdots+\alpha_r x_r), \tag{14}$$

确定系数 $\alpha_1, \alpha_2, \cdots, \alpha_r$ 使

$$(u, x_j)=\left(y_{r+1}-\sum_{i=1}^{r}\alpha_i x_i, \ x_j\right)=0, \quad j=1, 2, \cdots, r.$$

根据 $x_1, \cdots, x_r$ 之间的直交性,这个方程可简化为

$$(\boldsymbol{y}_{r+1}, \boldsymbol{x}_j)-\alpha_j=0,$$

即

$$\alpha_j=(\boldsymbol{y}_{r+1}, \boldsymbol{x}_j), \quad j=1, \cdots, r. \tag{15}$$

向量 $\boldsymbol{u}$ 显然不是零向量，因为不然的话，$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_r, \boldsymbol{y}_{r+1}$ 便为线性相关，从而 $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{r+1}$ 也为线性相关.

令

$$\boldsymbol{x}_{r+1}=\boldsymbol{u}/\|\boldsymbol{u}\|, \tag{16}$$

于是便求得了 $\boldsymbol{x}_{r+1}$. 继续此法，最后便得到标准直交系

$$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n.$$ 证毕.

定理中所用的方法，便是著名的 Gram-Schmidt 直交化方法.

由上述方法得到的直交系 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$，因为是线性独立的，其个数等于空间 $\mathscr{C}^n$ 的维数，所以可作为空间的一个基底.

定理 7 设 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ 是 $\mathscr{C}^n$ 中的一个线性独立系，若向量 $\boldsymbol{x}$ 和它们都直交，即

$$(\boldsymbol{x}, \boldsymbol{x}_i)=0, \quad i=1, 2, \cdots, n,$$

则必有 $\boldsymbol{x}=\mathbf{0}$.

证明 因为 $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ 构成空间 $\mathscr{C}^n$ 的一个基底，所以任何 $\boldsymbol{x}$ 都可以表示成它们的线性组合

$$\boldsymbol{x}=\alpha_1\boldsymbol{x}_1+\cdots+\alpha_n\boldsymbol{x}_n.$$

根据内积的性质，

$$(\boldsymbol{x}, \boldsymbol{x})=\left(\boldsymbol{x}, \sum_{i=1}^{n}\alpha_i\boldsymbol{x}_i\right)=\sum_{i=1}^{n}\bar{\alpha}_i(\boldsymbol{x}, \boldsymbol{x}_i)=0,$$

从而可知 $\boldsymbol{x}=\mathbf{0}$. 证毕.

3-3 内积的表示

前面已经讲过，在空间中建立基底以后，向量 $\boldsymbol{x}$ 可以用它的坐标 $\xi_1, \cdots, \xi_n$ 表示出来. 现在我们进一步利用空间的

基底和向量的坐标，把任意两个向量 x 和 y 的内积 (x, y) 表示出来. 为此我们假定已知 $\mathscr{C}^n$ 的一个基底为

$$q_1, q_2, \cdots, q_n,$$

它不一定是直交系，在这个基底下，向量 x 和 y 分别表示为

$$x=(\xi_1, \xi_2, \cdots, \xi_n), \quad y=(\eta_1, \eta_2, \cdots, \eta_n),$$

即

$$x=\xi_1 q_1+\xi_2 q_2+\cdots+\xi_n q_n,$$
$$y=\eta_1 q_1+\eta_2 q_2+\cdots+\eta_n q_n.$$

根据内积的性质，我们有关系式

$$(x, y)=\left(\sum_{i=1}^n \xi_i q_i, \sum_{j=1}^n \eta_j q_j\right)=\sum_{i,j=1}^n \xi_i \bar{\eta}_j (q_i, q_j)$$
$$=\sum_{i,j=1}^n h_{ij} \xi_i \bar{\eta}_j, \tag{17}$$

其中

$$h_{ij}=(q_i, q_j), \quad i, j=1, 2, \cdots, n \tag{18}$$

显然有

$$h_{ij}=\bar{h}_{ji}, \tag{19}$$

并且 h_{ij} 只与基底有关，而和向量 x 和 y 无关.

由此可知，酉空间 $\mathscr{C}^n$ 中的向量 x 和 y 的内积，可以表示成一个二次型(17)，其矩阵为

$$H=[h_{ij}], \tag{20}$$

据(19)知，$H=H^*$，$*$ 表示共轭转置，这就是说，H 是一个 Hermite 矩阵.

其次，因为 $(x, x)\geqslant 0$，只有 $x=0$ 时才有 $(x, x)=0$，所以二次形(17)为正定，即，H 是一个正定矩阵.

于特例，若取空间基底为标准直交基底

$$e_1, e_2, \cdots, e_n,$$

因为 $(e_i, e_j)=\delta_{ij}$, $i, j=1, \cdots, n$，这时，内积的表达式(17)便化为非常简单的形式：

$$(\boldsymbol{x}, \boldsymbol{y})=\sum_{i=1}^{n} \xi_i \bar{\eta}_i. \tag{21}$$

同理可知，欧氏空间中的内积可以表示成一个实二次型，其矩阵 A 为对称正定，并且(21)化为

$$(\boldsymbol{x}, \boldsymbol{y})=\sum_{i=1}^{n} \xi_i \eta_i. \tag{22}$$

如果把 $\boldsymbol{x}$ 在空间基底下的坐标向量$(\xi_1, \xi_2, \cdots, \xi_n)$表示成列向量

$$\boldsymbol{x}=\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}, \tag{23}$$

不难证明，对空间 $\mathscr{C}^n$，

$$(\boldsymbol{x}, \boldsymbol{x})=\boldsymbol{x}^* H \boldsymbol{x}, \tag{24}$$

对欧氏空间 $\mathscr{R}^n$，

$$(\boldsymbol{x}, \boldsymbol{x})=\boldsymbol{x}^T A \boldsymbol{x}, \tag{25}$$

其中的 A 是一个对称正定矩阵．若基底为标准直交系，则(24)和(25)分别化为

$$\|\boldsymbol{x}\|^2=\sum_{i=1}^{n} |\xi_i|^2,$$

$$\|\boldsymbol{x}\|^2=\sum_{i=1}^{n} \xi_i^2.$$

3-4　空间的直交分解

前面已经讲过空间的直接和分解，有了内积，就可以进一步考虑一种特殊的直接和分解，为定义直交投影奠定基础．今设 $\mathscr{W}$ 是酉空间 $\mathscr{C}^n$ 的一个 k 维子空间，并设 $\boldsymbol{q}_1, \boldsymbol{q}_2, \cdots, \boldsymbol{q}_k$ 是它的一个直交基底．对于整个空间 $\mathscr{C}^n$ 来说．我们可以再补充 $n-k$ 个直交向量 $\boldsymbol{q}_{k+1}, \cdots, \boldsymbol{q}_n$，使 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n$ 构成 $\mathscr{C}^n$ 的一个直交基底．记 $\boldsymbol{q}_{k+1}, \cdots, \boldsymbol{q}_n$ 所张的子空间为

$$\mathrm{Span}\{\boldsymbol{q}_{k+1}, \cdots, \boldsymbol{q}_n\},$$

则其维数为$n-k$，并且，其中的任一向量$\boldsymbol{x}$和子空间$\mathscr{W}$的任何向量都直交，因此，便称这个$n-k$维子空间为$\mathscr{W}$的直交补空间，记作

$$\mathscr{W}^{\perp}=\mathrm{Span}\{\boldsymbol{q}_{k+1}, \cdots, \boldsymbol{q}_n\}.$$

于是便得到空间$\mathscr{C}^n$的一种直接和分解

$$\mathscr{C}^n=\mathscr{W}\oplus\mathscr{W}^{\perp}. \tag{26}$$

空间$\mathscr{C}^n$还可以进一步分解成若干个互相直交的子空间的直接和．于特例，对于给定的直交基底$\{\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n\}$，$\mathscr{C}^n$可以分解成直接和

$$\mathscr{C}^n=\mathrm{Span}\{\boldsymbol{q}_1\}\oplus\mathrm{Span}\{\boldsymbol{q}_2\}\oplus\cdots\oplus\mathrm{Span}\{\boldsymbol{q}_n\}. \tag{27}$$

这就是说，$\mathscr{C}^n$可以分解成n个一维直交子空间的直接和．

欧氏空间$\mathscr{R}^n$也可以作同样的直交分解，这里不再重复．

§4 线性算子及其矩阵表示

4-1 线性算子的定义

设$\mathscr{V}_1$和$\mathscr{V}_2$是同一数域K上的两个向量空间，$\boldsymbol{A}$是从$\mathscr{V}_1$到$\mathscr{V}_2$的一个映射，即对任何$\boldsymbol{x}\in\mathscr{V}_1$,

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}\in\mathscr{V}_2,$$

若映射$\boldsymbol{A}$具有下列性质：

1° 对任意的$\boldsymbol{x}\in\mathscr{V}_1$, $\alpha\in K$, $\boldsymbol{A}(\alpha\boldsymbol{x})=\alpha\boldsymbol{A}\boldsymbol{x}$,

2° 对任意的$\boldsymbol{x}$、$\boldsymbol{y}\in\mathscr{V}_1$, $\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{y}$，便称映射$\boldsymbol{A}$为从向量空间$\mathscr{V}_1$到向量空间$\mathscr{V}_2$内的一个**线性算子**，并称$\mathscr{V}_1$为$\boldsymbol{A}$的定义域，称集合

$$\boldsymbol{A}\mathscr{V}_1\subset\mathscr{V}_2$$

为算子$\boldsymbol{A}$的**像空间**，常记成$\mathscr{R}(\boldsymbol{A})$．

因为，对任何 $\boldsymbol{x}\in\mathscr{V}_1$，都有

$$\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{0})=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}\boldsymbol{0}=\boldsymbol{A}\boldsymbol{x}.$$

所以

$$\boldsymbol{A}\boldsymbol{0}=\boldsymbol{0},$$

即 $\mathscr{V}_1$ 中的零向量在 $\mathscr{V}_2$ 中的像仍为零向量.

线性算子 $\boldsymbol{A}$ 的像空间 $\mathscr{R}(\boldsymbol{A})$ 必为 $\mathscr{V}_2$ 内的一个子空间. 因为，若

$$\boldsymbol{u},\ \boldsymbol{v}\in\mathscr{R}(\boldsymbol{A}),$$

则必存在 $\boldsymbol{x},\ \boldsymbol{y}\in\mathscr{V}_1$，使

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{u},\ \boldsymbol{A}\boldsymbol{y}=\boldsymbol{v}.$$

于是有，对任意的 $\alpha,\ \beta\in K$，

$$\boldsymbol{A}(\alpha\boldsymbol{x}+\beta\boldsymbol{y})=\alpha\boldsymbol{A}\boldsymbol{x}+\beta\boldsymbol{A}\boldsymbol{y}=\alpha\boldsymbol{u}+\beta\boldsymbol{v},$$

所以

$$\alpha\boldsymbol{u}+\beta\boldsymbol{v}\in\boldsymbol{A}\mathscr{V}_1,$$

由此可见 $\mathscr{R}(\boldsymbol{A})$ 是 $\mathscr{V}_2$ 内的一个子空间.

如果对一切 $\boldsymbol{x}\in\mathscr{V}_1$，恒有

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{0},$$

便称 $\boldsymbol{A}$ 是零算子，常记之为 $\boldsymbol{0}$.

算子的运算

(i) $\alpha\boldsymbol{A}$　设 $\alpha\in K$，今规定 $\boldsymbol{B}=\alpha\boldsymbol{A}$.

若 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}$，则

$$\boldsymbol{B}\boldsymbol{x}=\alpha\boldsymbol{A}\boldsymbol{x}=\alpha\boldsymbol{y}. \tag{28}$$

(ii) $\boldsymbol{A}+\boldsymbol{B}$　设 $\boldsymbol{B}$ 也是一个从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子，今规定：对 $\boldsymbol{x}\in\mathscr{V}_1$，

$$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{B}\boldsymbol{x}, \tag{29}$$

并称 $\boldsymbol{C}=\boldsymbol{A}+\boldsymbol{B}$ 为算子 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的和算子.

(iii) $\boldsymbol{B}\boldsymbol{A}$　设 $\boldsymbol{A}$ 为从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$，$\boldsymbol{B}$ 为从 $\mathscr{V}_2$ 到 $\mathscr{V}_3$ 内的线性算子，$\mathscr{V}_1$、$\mathscr{V}_2$、$\mathscr{V}_3$ 为同一数域 K 上的向量空间. 对 $\boldsymbol{x}\in\mathscr{V}_1$，令

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{z}=\boldsymbol{B}\boldsymbol{y},$$

我们规定

$$(\boldsymbol{BA})\boldsymbol{x}=\boldsymbol{B}(\boldsymbol{Ax})=\boldsymbol{By}=\boldsymbol{z}, \tag{30}$$

称 $\boldsymbol{BA}$ 为算子 $\boldsymbol{B}$ 和 $\boldsymbol{A}$ 的乘积.

容易证明，上述 $\alpha\boldsymbol{A}$, $\boldsymbol{A}+\boldsymbol{B}$, $\boldsymbol{BA}$ 仍是线性算子，但后者的像空间在 $\mathscr{V}_3$ 内，$\mathscr{R}(\boldsymbol{BA})\subset\mathscr{V}_3$.

4-2　线性算子的零空间和算子的秩

设 $\boldsymbol{A}$ 是从向量空间 $\mathscr{V}_1$ 到同一数域 K 上的向量空间 $\mathscr{V}_2$ 内的线性算子.

定义 10　在空间 $\mathscr{V}_1$ 中满足方程

$$\boldsymbol{Ax}=\boldsymbol{0} \tag{31}$$

的一切向量 $\boldsymbol{x}$ 所组成的集合称为线性算子 $\boldsymbol{A}$ 的零空间，记作 $\mathscr{N}(\boldsymbol{A})$.

显然，$\mathscr{N}(\boldsymbol{A})$是 $\mathscr{V}_1$ 的一个子空间，因为，若

$$\boldsymbol{Ax}=\boldsymbol{0},\ \boldsymbol{Ay}=0,$$

则对任何 α, $\beta\in K$ 必有

$$\boldsymbol{A}(\alpha\boldsymbol{x}+\beta\boldsymbol{y})=\alpha\boldsymbol{Ax}+\beta\boldsymbol{Ay}=\boldsymbol{0},$$

即

$$\alpha\boldsymbol{x}+\beta\boldsymbol{y}\in\mathscr{N}(\boldsymbol{A}).$$

应当注意，算子 $\boldsymbol{A}$ 的像空间 $\mathscr{R}(\boldsymbol{A})$是向量空间 $\mathscr{V}_2$ 内的一个子空间，而零空间 $\mathscr{N}(\boldsymbol{A})$则为向量空间 $\mathscr{V}_1$ 内的一个子空间.

定义 11　我们规定像空间 $\mathscr{R}(\boldsymbol{A})$的维数为线性算子的秩，记作

$$\mathrm{rank}(\boldsymbol{A})=\dim\mathscr{R}(\boldsymbol{A}). \tag{32}$$

像空间 $\mathscr{R}(\boldsymbol{A})$，零空间 $\mathscr{N}(\boldsymbol{A})$的维数和向量空间 $\mathscr{V}_1$ 的维数之间存在有以下的关系：

定理 8

$$\dim\mathscr{V}_1=\dim\mathscr{R}(\boldsymbol{A})+\dim\mathscr{N}(\boldsymbol{A}). \tag{33}$$

证明　设　$\dim \mathscr{V}_1=n$, $\dim \mathscr{N}(\boldsymbol{A})=r$.
$\mathscr{N}(\boldsymbol{A})$中必存在基底$\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_r$. 即,

$$\boldsymbol{A}\boldsymbol{p}_i=\boldsymbol{0}, \quad i=1, 2, \cdots, r,$$

且$\{\boldsymbol{p}_1, \cdots, \boldsymbol{p}_r\}$为线性独立. 因为$\dim \mathscr{V}_1=n$, 所以可在$\mathscr{N}(\boldsymbol{A})$之外补充$n-r$个向量$\boldsymbol{p}_{r+1}, \cdots, \boldsymbol{p}_n$, 使

$$\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$$

构成$\mathscr{V}_1$的一个基底. 令

$$\boldsymbol{q}_j=\boldsymbol{A}\boldsymbol{p}_j, \quad j=r+1, \cdots, n.$$

显然有:

1) 因为$\boldsymbol{p}_{r+1}, \cdots, \boldsymbol{p}_n$不属于$\mathscr{N}(\boldsymbol{A})$, 所以

$$\boldsymbol{q}_j=\boldsymbol{A}\boldsymbol{p}_j\neq\boldsymbol{0}, \quad j=r+1, \cdots, n.$$

2) $\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$为线性独立. 因为, 否则的话, 必存在关系

$$\alpha_{r+1}\boldsymbol{A}\boldsymbol{p}_{r+1}+\cdots+\alpha_n\boldsymbol{A}\boldsymbol{p}_n=\boldsymbol{0},$$

即
$$\boldsymbol{A}(\alpha_{r+1}\boldsymbol{p}_{r+1}+\cdots+\alpha_n\boldsymbol{p}_n)=\boldsymbol{0},$$

从而应有
$$\alpha_{r+1}\boldsymbol{p}_{r+1}+\cdots+\alpha_n\boldsymbol{p}_n\in\mathscr{N}(\boldsymbol{A}).$$

除$\alpha_{r+1}=\alpha_{r+2}=\cdots=\alpha_n=0$外, 这是不可能的. 否则的话, 便意味着$\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$为线性相关.

现在我们来证明

$$\dim \mathscr{R}(\boldsymbol{A})=n-r.$$

为此, 只需证明$\mathscr{R}(\boldsymbol{A})$中的任何向量$\boldsymbol{y}$都可以表示成$\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$的线性组合即可. 由于$\boldsymbol{y}\in\mathscr{R}(\boldsymbol{A})$, 所以存在有向量$\boldsymbol{x}\in\mathscr{V}_1$, 使

$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}.$$

设
$$\boldsymbol{x}=\beta_1\boldsymbol{p}_1+\cdots+\beta_n\boldsymbol{p}_n,$$

则
$$\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}\left(\sum_{i=1}^{n}\beta_i\boldsymbol{p}_i\right)=\beta_{r+1}\boldsymbol{q}_{r+1}+\cdots+\beta_n\boldsymbol{q}_n,$$

根据$\boldsymbol{q}_{r+1}, \cdots, \boldsymbol{q}_n$的线性独立性, 这种表示是唯一的, 这就证

明了关系 $\dim \mathscr{R}(\boldsymbol{A})=n-r$. 证毕.

推论 向量空间 $\mathscr{V}_1$ 可以作直接和分解:

$$\mathscr{V}_1=\mathscr{N}(\boldsymbol{A})\oplus \operatorname{Span}\{\boldsymbol{p}_{r+1}, \cdots, \boldsymbol{p}_n\}. \tag{34}$$

4-3 线性算子的矩阵表示

在向量空间中建立了坐标系以后，空间中的向量都可以表示为坐标向量. 如果在向量空间 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 中都建立了坐标系，那末，从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子 $\boldsymbol{A}$ 也可以给出一种具体的表示. 我们先假定

$$\dim \mathscr{V}_1=n, \quad \dim \mathscr{V}_2=m.$$

在 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 中分别取定基底:

$$\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n \text{ 和 } \boldsymbol{g}_1, \cdots, \boldsymbol{g}_m,$$

因为 $\boldsymbol{A}\boldsymbol{h}_j\in \mathscr{V}_2$, $j=1, 2, \cdots, n$, 所以它们都可以唯一地表示成 $\boldsymbol{g}_1, \cdots, \boldsymbol{g}_m$ 的线性组合:

$$\boldsymbol{A}\boldsymbol{h}_j=\sum_{i=1}^{m} a_{ij}\boldsymbol{g}_i, \quad j=1, \cdots, n, \tag{35}$$

次设 $\boldsymbol{x}\in \mathscr{V}_1$, 并假定它在基底 $\{\boldsymbol{h}_1, \cdots, \boldsymbol{h}_n\}$ 下的表示为:

$$\boldsymbol{x}=\xi_1\boldsymbol{h}_1+\xi_2\boldsymbol{h}_2+\cdots+\xi_n\boldsymbol{h}_n, \tag{36}$$

令 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$, 向量 $\boldsymbol{y}$ 可以用基底 $\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_m$ 表示成

$$\boldsymbol{y}=\eta_1\boldsymbol{g}_1+\eta_2\boldsymbol{g}_2+\cdots+\eta_m\boldsymbol{g}_m. \tag{37}$$

用 ξ_j 分别乘(35)两端, 并对 j 求和得

$$\begin{aligned}\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}&=\boldsymbol{A}\left(\sum_{j=1}^{n}\xi_j\boldsymbol{h}_j\right)=\sum_{j=1}^{n}\xi_j\boldsymbol{A}\boldsymbol{h}_j\\&=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}a_{ij}\xi_j\right)\boldsymbol{g}_i=\sum_{i=1}^{m}\eta_i\boldsymbol{g}_i,\end{aligned}$$

于是便得到了向量 $\boldsymbol{y}$ 的坐标 $(\eta_1, \cdots, \eta_m)$ 和向量 $\boldsymbol{x}$ 的坐标 $(\xi_1, \xi_2, \cdots, \xi_n)$ 之间的关系

$$\begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}. \tag{38}$$

这就是说，在给定的空间基底下，线性算子 $\boldsymbol{A}$ 可以唯一地表示成 $m\times n$ 阶矩阵$[a_{ij}]$，其元素 a_{ij} 都是数域 K 中的数。若分别用 $\boldsymbol{x}_h$, $\boldsymbol{y}_g$ 表示向量在有关基底下的表示，记 $A_{h,g}$ 为线性算子 $\boldsymbol{A}$ 在有关基底下的矩阵表示，则(38)可简写成：

$$\boldsymbol{y}_g = A_{h,g}\boldsymbol{x}_h. \tag{39}$$

反之，在向量空间 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 中分别给定基底$\{\boldsymbol{h}_j\}$和$\{\boldsymbol{g}_i\}$以后，任何一个元素属于域 K 的 $m\times n$ 阶矩阵 $[a_{ij}]$，由对应关系(35)确定了一个唯一的从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子 $\boldsymbol{A}$。这就是说，在给定空间基底的情况下，从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 的线性算子 $\boldsymbol{A}$ 和 $m\times n$ 阶矩阵$[a_{ij}]$为一一对应。

如果在空间 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 中，我们选取另外两组基底

$$\boldsymbol{h}'_1,\ \boldsymbol{h}'_2,\ \cdots,\ \boldsymbol{h}'_n \text{ 和 } \boldsymbol{g}'_1,\ \boldsymbol{g}'_2,\ \cdots,\ \boldsymbol{g}'_m,$$

并记 $\boldsymbol{x}'_h$, $\boldsymbol{y}'_g$, $A_{h',g'}$ 分别为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 以及线性算子 $\boldsymbol{A}$ 在有关基底下的表示，则

$$\boldsymbol{y}_{g'} = A_{h',g'}\boldsymbol{x}_{h'}. \tag{40}$$

现在我们进一步来研究同一线性算子 $\boldsymbol{A}$ 在两种不同坐标系中的矩阵表示之间的关系。

今假定空间 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 的新、旧两个基底之间的关系分别为

$$\boldsymbol{h}_j = \sum_{i=1}^{n} c_{ij}\boldsymbol{h}'_i \quad j=1,\ 2,\ \cdots,\ n,$$

$$\boldsymbol{g}'_i = \sum_{j=1}^{m} b_{ij}\boldsymbol{g}_j \quad i=1,\ 2,\ \cdots,\ m,$$

并设向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 在新坐标系中的坐标向量分别为

$$(\xi_1', \xi_2', \cdots, \xi_n') \quad 和 \quad (\eta_1', \eta_2', \cdots, \eta_m'),$$

即 $$\boldsymbol{x}=\xi_1'\boldsymbol{h}_1'+\cdots+\xi_n'\boldsymbol{h}_n', \quad \boldsymbol{y}=\eta_1'\boldsymbol{g}_1'+\cdots+\eta_m'\boldsymbol{g}_m'.$$

令 $C=[c_{ij}]$, $B=[b_{ij}]$, 于是便得到向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 在新旧坐标下的坐标向量之间的关系:

$$\begin{aligned} \boldsymbol{x}_{h'} &= C\boldsymbol{x}_h \\ \boldsymbol{y}_g &= B\boldsymbol{y}_{g'} \end{aligned} \tag{41}$$

因为 $$\boldsymbol{y}_{g'}=A_{h',g'}\boldsymbol{x}_{h'}=A_{h',g'}C\boldsymbol{x}_h,$$

所以

$$\boldsymbol{y}_g=B\boldsymbol{y}_{g'}=BA_{h',g'}C\boldsymbol{x}_h. \tag{42}$$

但已知 $$\boldsymbol{y}_g=A_{h,g}\boldsymbol{x}_h,$$

所以有

$$A_{h,g}=BA_{h',g'}C \tag{43}$$

这就是线性算子 $\boldsymbol{A}$ 在不同基底下的矩阵表示之间的关系，其中的 C 和 B 分别为 $n\times n$ 阶和 $m\times m$ 阶方阵. 由于坐标基底的线性独立性，可以证明它们都是非奇异的. 我们称(41)为坐标变换公式，称 C 和 B 为确定相应线性变换的矩阵.

定义 12 设 A 和 $\widetilde{A}$ 都是 $m\times n$ 阶矩阵，若存在非奇异 $m\times m$ 阶和 $n\times n$ 阶方阵 B 和 C 使

$$\widetilde{A}=BAC, \tag{44}$$

则称矩阵 A 和 $\widetilde{A}$ 为等价.

从以上的讨论可以看出，所谓等价矩阵，实际上就是同一线性算子在不同空间基底下的表示.

象空间的直接和分解及线性算子在子空间上的表示.

设 $\mathrm{rank}(A)=r$, 并假定按某种方法对 $\mathscr{R}(\boldsymbol{A})$ 作直接和分解

$$\mathscr{R}(\boldsymbol{A})=\mathscr{R}_1\oplus\mathscr{R}_2\oplus\cdots\oplus\mathscr{R}_k, \tag{45}$$

其中，$\dim \mathscr{R}_i = r_i \geqslant 1$，$\sum_{i=1}^{k} \dim \mathscr{R}_i = r$. 于是显然应有 $k \leqslant r$. 记空间 $\mathscr{V}_1$ 中与 $\mathscr{R}_i$ 相应的向量集合为 S_i，$i=1, 2, \cdots, k$，则容易证明 S_i 都是子空间，并且有直接和分解：

$$\mathscr{V}_1 = S_1 \oplus \cdots \oplus S_k \oplus \mathscr{N}(\boldsymbol{A}). \tag{46}$$

今对 $i=1, 2, \cdots, k$ 规定算子 $\boldsymbol{A}_i$ 如下：

$$\boldsymbol{A}_i \boldsymbol{u} = \boldsymbol{A}_i(\boldsymbol{u}_1 + \cdots + \boldsymbol{u}_{k+1}) = \boldsymbol{A}\boldsymbol{u}_i,$$

其中，　　$\boldsymbol{u}_i \in S_i$，　$i=1, \cdots, k$，　$\boldsymbol{u}_{k+1} \in \mathscr{N}(\boldsymbol{A})$.

容易证明，$\boldsymbol{A}_i$ 都是线性算子，并且它们的和就是 $\boldsymbol{A}$，即，

$$\boldsymbol{A} = \boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_k. \tag{48}$$

在 S_i 和 $\mathscr{R}_i$ 中分别建立坐标系以后，记 $\boldsymbol{A}_i$ 的矩阵表示为 A_i，它是一个 $r_i \times r_i$ 阶非奇异方阵，而算子 $\boldsymbol{A}$ 在子空间

$$S_1 \oplus S_2 \oplus \cdots \oplus S_k$$

上的矩阵表示为

$$\begin{bmatrix} A_1 & & & \\ & A_2 & & 0 \\ 0 & & \ddots & \\ & & & A_k \end{bmatrix}. \tag{49}$$

如果在 $\mathscr{N}(\boldsymbol{A})$ 中也选取一个基底，不论这个基底如何选取，算子 $\boldsymbol{A}$ 在整个空间 $\mathscr{V}_1$ 上关于上述基底下的矩阵表示为

$$A = \left[\begin{array}{ccc|c} A_1 & & 0 & \\ 0 & \ddots & & 0 \\ & & A_k & \\ \hline & 0 & & 0 \end{array}\right] \tag{50}$$

这里的 A 是一个 $m \times n$ 阶矩阵. 应当指出，在空间 $\mathscr{V}_2$ 中也

要补充 $m-r$ 个向量以建立基底，而上述矩阵表示则与补充的向量系无关.

作为特例，如果在 $\mathscr{R}_i$ 和 S_i 中所取基底

$$\boldsymbol{q}_{k_1},\ \boldsymbol{q}_{k_2},\ \cdots,\ \boldsymbol{q}_{kr_i} \text{ 和 } \boldsymbol{p}_{k_1},\ \boldsymbol{p}_{k_2},\ \cdots,\ \boldsymbol{p}_{kr_i}$$

之间有关系 $\quad \boldsymbol{q}_{k_j}=\boldsymbol{A}\boldsymbol{p}_{k_j},\quad j=1,\ 2,\ \cdots,\ r_i,$

则算子 $\boldsymbol{A}_i$ 的矩阵表示 A_i 就是 $r_i\times r_i$ 阶单位阵 I_{r_i}，从而 $\boldsymbol{A}$ 的矩阵表示为

$$A=\left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right]. \tag{51}$$

由此不难看出，一切秩为 r 的 $m\times n$ 阶矩阵都和(51)中右端的矩阵等价.

§5 酉空间上的线性算子

5-1 酉空间中向量的表示

因为酉空间中定义有内积，可以用它把坐标向量表示出来. 如取 n 维酉空间 $\mathscr{C}^n$ 中的一个标准直交系

$$\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \cdots,\ \boldsymbol{e}_n$$

作为基底，向量 $\boldsymbol{x}$ 在这个基底下的表示为

$$\boldsymbol{x}=\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{bmatrix},$$

即，

$$\boldsymbol{x}=\xi_1\boldsymbol{e}_1+\xi_2\boldsymbol{e}_2+\cdots+\xi_n\boldsymbol{e}_n. \tag{52}$$

因为 $\{\boldsymbol{e}_i\}$ 是一个标准直交系，所以

$$(\boldsymbol{x},\ \boldsymbol{e}_j)=\sum_{i=1}^{n}\xi_j(\boldsymbol{e}_i,\ \boldsymbol{e}_j)=\xi_j,$$

于是便得到向量 $\boldsymbol{x}$ 的表示式

$$\boldsymbol{x}=\begin{bmatrix}(\boldsymbol{x},\ \boldsymbol{e}_1)\\(\boldsymbol{x},\ \boldsymbol{e}_2)\\\vdots\\(\boldsymbol{x},\ \boldsymbol{e}_n)\end{bmatrix}. \tag{53}$$

而向量 $\boldsymbol{x}$ 的长度为

$$|\boldsymbol{x}|^2=\sum_{i=1}^{n}|\xi_i|^2.$$

5-2 伴随算子

设 $\boldsymbol{A}$ 是从酉空间 $\mathscr{C}^n\to\mathscr{C}^n$ 内的一个线性算子，$\boldsymbol{A}^*$ 也是一个从 $\mathscr{C}^n\to\mathscr{C}^n$ 内的线性算子，如果对任意两个向量 $\boldsymbol{x},\ \boldsymbol{y}\in\mathscr{C}^n$，恒有

$$(\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{y})=(\boldsymbol{x},\ \boldsymbol{A}^*\boldsymbol{y}), \tag{54}$$

便称 $\boldsymbol{A}^*$ 为 $\boldsymbol{A}$ 的伴随算子．因为，如果(54)成立，则必有

$$(\boldsymbol{A}^*\boldsymbol{y},\ \boldsymbol{x})=(\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{x}),$$

所以，$\boldsymbol{A}$ 也是 $\boldsymbol{A}^*$ 的伴随算子．

定理 9 对于任何线性算子 $\boldsymbol{A}(\mathscr{C}^n\to\mathscr{C}^n)$，恒存在唯一的伴随算子 $\boldsymbol{A}^*$.

证明 在酉空间 $\mathscr{C}^n$ 中取标准直交基底

$$\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ \cdots,\ \boldsymbol{e}_n,$$

$\boldsymbol{A}$ 的伴随算子 $\boldsymbol{A}^*$ 如果存在的话，对任何向量 $\boldsymbol{y}\in\mathscr{C}^n$，它应满足关系

$$\boldsymbol{A}^*\boldsymbol{y}=\sum_{i=1}^{n}(\boldsymbol{A}^*\boldsymbol{y},\ \boldsymbol{e}_i)\boldsymbol{e}_i$$

即

$$\boldsymbol{A}^*\boldsymbol{y}=\sum_{i=1}^{n}(\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i. \tag{55}$$

(55)式的右端不显含 $\boldsymbol{A}^*$，所以这个方程便可以作为 $\boldsymbol{A}$ 的伴随算子 $\boldsymbol{A}^*$ 的定义，满足(55)的算子 $\boldsymbol{A}^*$ 也必满足(54)．

现在我们来证明，由(55)定义的算子 $\boldsymbol{A}^*$ 是线性算子. 对任意的复数 $\alpha\in\mathscr{C}$，向量 $\boldsymbol{y}\in\mathscr{C}^n$，

$$\boldsymbol{A}^*(\alpha\boldsymbol{y})=\sum_{i=1}^{n}(\alpha\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i=\alpha\sum_{i=1}^{n}(\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i=\alpha\boldsymbol{A}^*\boldsymbol{y},$$

所以，$\boldsymbol{A}^*$ 具有齐次性. 次设 $\boldsymbol{x}$、$\boldsymbol{y}$ 为 $\mathscr{C}^n$ 中任意两个向量，

$$\boldsymbol{A}^*(\boldsymbol{x}+\boldsymbol{y})=\sum_{i=1}^{n}(\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i=\sum_{i=1}^{n}(\boldsymbol{x},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i$$

$$+\sum_{i=1}^{n}(\boldsymbol{y},\ \boldsymbol{A}\boldsymbol{e}_i)\boldsymbol{e}_i=\boldsymbol{A}^*\boldsymbol{x}+\boldsymbol{A}^*\boldsymbol{y},$$

所以，$\boldsymbol{A}^*$ 也具有可加性，从而可知，$\boldsymbol{A}^*$ 是一个线性算子.

关于唯一性问题，可设 $\boldsymbol{B}$ 也是由(55)所定义的算子，则对任意的 $\boldsymbol{y}\in\mathscr{C}^n$，$\boldsymbol{A}^*\boldsymbol{y}=\boldsymbol{B}\boldsymbol{y}$，即

$$(\boldsymbol{A}^*-\boldsymbol{B})\boldsymbol{y}=\boldsymbol{0},\quad 对一切\ \boldsymbol{y}\in\mathscr{C}^n.$$

从而必有 $\boldsymbol{A}^*-\boldsymbol{B}=\boldsymbol{0}$，即

$$\boldsymbol{B}=\boldsymbol{A}^*.\ 证毕.$$

这里顺便指出，伴随算子的概念，可以推广到从 $\mathscr{C}^n\to\mathscr{C}^m$ 的线性算子上来.

5-3 伴随算子的矩阵表示

今假定在酉空间 $\mathscr{C}^n$ 中取定了一个标准直交系 $\{\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n\}$，以其作为基底，从 $\mathscr{C}^n\to\mathscr{C}^n$ 的线性算子 $\boldsymbol{A}$ 的矩阵表示为

$$A=[a_{ij}],$$

它是一个 $n\times n$ 阶的复矩阵. 记 A 的 n 个列向量为 $\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n$，则有

$$A\boldsymbol{e}_k=\boldsymbol{a}_k=\sum_{i=1}^{n}(\boldsymbol{a}_k,\ \boldsymbol{e}_i)\boldsymbol{e}_i=\sum_{i=1}^{n}a_{ik}\boldsymbol{e}_i=\sum_{i=1}^{n}(A\boldsymbol{e}_k,\ \boldsymbol{e}_i)\boldsymbol{e}_i,$$

所以 $\boldsymbol{a}_k$ 的第 i 个分量

$$a_{ik}=(A\boldsymbol{e}_k,\ \boldsymbol{e}_i),\quad i、k=1,\ 2,\ \cdots,\ n.$$

在同一坐标系中，设伴随算子 $\boldsymbol{A}^*$ 的矩阵表示为 $[a_{ik}^*]$，

于是

$$a_{ik}^* = (\boldsymbol{A}^* \boldsymbol{e}_k,\ \boldsymbol{e}_i),\quad i、k=1,\ 2,\ \cdots,\ n.$$

从而可知

$$a_{ik}^* = \bar{a}_{ki},\quad i、k=1,\ 2,\ \cdots,\ n.$$

即

$$[a_{ik}^*] = [\bar{a}_{ik}]^T. \tag{56}$$

这说明，$\boldsymbol{A}$ 的伴随算子 $\boldsymbol{A}^*$ 的矩阵表示为 $\boldsymbol{A}$ 的矩阵表示的共轭转置.

伴随算子的一些简单性质 根据伴随算子的定义，不难证明下列性质:

1) $(\boldsymbol{A}^*)^* = \boldsymbol{A}$,

2) $(\boldsymbol{A}+\boldsymbol{B})^* = \boldsymbol{A}^* + \boldsymbol{B}^*$,

3) $(\alpha \boldsymbol{A})^* = \bar{\alpha} \boldsymbol{A}^*$,

4) $(\boldsymbol{A}\boldsymbol{B})^* = \boldsymbol{B}^* \boldsymbol{A}^*$.

关于伴随算子的矩阵表示以及上列简单性质，都不难推广到从 $\mathscr{C}^n \to \mathscr{C}^m$ 的线性算子.

定理 10 设 $\boldsymbol{A}$ 为从西空间 $\mathscr{C}^n \to \mathscr{C}^m$ 的线性算子，$\boldsymbol{A}^*$ 是它的伴随算子，则

$$\mathscr{R}(\boldsymbol{A}^*) = \mathscr{N}^{\perp}(\boldsymbol{A}). \tag{57}$$

证明 设 $\boldsymbol{x} \in \mathscr{C}^n$, $\boldsymbol{y} \in \mathscr{C}^m$, $\boldsymbol{x}$、$\boldsymbol{y}$ 为任意，则

$$(\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{y}) = (\boldsymbol{x},\ \boldsymbol{A}^*\boldsymbol{y}).$$

若 $\boldsymbol{x} \in \mathscr{N}(\boldsymbol{A})$，则显然有

$$(\boldsymbol{x},\ \boldsymbol{A}^*\boldsymbol{y}) = 0,$$

由于 $\boldsymbol{y}$ 为任意，所以 $\mathscr{N}(\boldsymbol{A}) \subset \mathscr{R}^{\perp}(\boldsymbol{A}^*)$.

反之，对给定的 $\boldsymbol{x} \in \mathscr{C}^n$，若对任意的 $\boldsymbol{y} \in \mathscr{C}^m$ 恒有

$$(\boldsymbol{x},\ \boldsymbol{A}^*\boldsymbol{y}) = 0,$$

即

$$\boldsymbol{x} \in \mathscr{R}^{\perp}(\boldsymbol{A}^*).$$

于特例，令 $\boldsymbol{y}=\boldsymbol{Ax}$ 便得

$$(\boldsymbol{Ax},\ \boldsymbol{Ax})=0,$$

所以必有 $\boldsymbol{Ax}=\boldsymbol{0}$，即 $\boldsymbol{x}\in\mathscr{N}(\boldsymbol{A})$.

因此有 $$\mathscr{R}(\boldsymbol{A}^*)=\mathscr{N}^{\perp}(\boldsymbol{A}).$$ 证毕.

定义 13 设 $\boldsymbol{A}$ 为从 $\mathscr{C}^n\to\mathscr{C}^n$ 的线性算子，$\boldsymbol{A}^*$ 是它的伴随算子，若

$$\boldsymbol{A}=\boldsymbol{A}^*, \tag{58}$$

则称 $\boldsymbol{A}$ 为**自伴算子**.

因为 $$(\boldsymbol{A}^*)^*=\boldsymbol{A},$$

所以，若 $\boldsymbol{A}$ 为自伴算子，则其伴随算子 $\boldsymbol{A}^*$ 也是自伴算子.

对于从 $\mathscr{C}^n\to\mathscr{C}^m$ 的线性算子 $\boldsymbol{A}$ 及其伴随算子 $\boldsymbol{A}^*$，不能直接考虑(58)，但算子积

$$\boldsymbol{AA}^* \text{ 和 } \boldsymbol{A}^*\boldsymbol{A}$$

分别是从 $\mathscr{C}^m\to\mathscr{C}^m$ 及从 $\mathscr{C}^n\to\mathscr{C}^n$ 的自伴算子.

根据伴随算子 $\boldsymbol{A}^*$ 的矩阵表示可知，若 $\boldsymbol{A}$ 为自伴，则

$$[a_{ij}]^*=[a_{ij}],$$

其中$[a_{ij}]$为 A 的矩阵表示，*表示转置共轭运算. 即，自伴算子的矩阵表示是 Hermite 矩阵.

对于欧氏空间上的线性算子 $\boldsymbol{A}$，其伴随算子 $\boldsymbol{A}^*$ 的矩阵表示为 $\boldsymbol{A}$ 的矩阵表示的转置，即

$$[\boldsymbol{A}^*]=[\boldsymbol{A}]^T,$$

$[A]$表示 $\boldsymbol{A}$ 的矩阵表示. 同理可知，欧氏空间上自伴算子的矩阵表示是实对称矩阵.

5-4 酉算子和直交算子

定义 14 设 $\boldsymbol{U}$ 是从酉空间 $\mathscr{C}^n\to\mathscr{C}^n$ 内的线性算子，若

$$\boldsymbol{UU}^*=\boldsymbol{U}^*\boldsymbol{U}=\boldsymbol{I}, \tag{59}$$

$\boldsymbol{I}$ 为单位算子，则称 $\boldsymbol{U}$ 是一个**酉算子**.

对于从欧氏空间 $\mathscr{R}^n \to \mathscr{R}^n$ 内的线性算子 $\boldsymbol{U}$，若满足关系(59)，则称它是一个**直交算子**.

定理11 从酉空间 $\mathscr{C}^n \to \mathscr{C}^n$ 内的线性算子 $\boldsymbol{U}$ 是酉算子的充分必要条件为，它保持向量的内积不变，即，对任意两个向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}^n$，恒有

$$(\boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y}). \tag{60}$$

证明 设 $\boldsymbol{U}$ 是一个酉算子，则

$$\boldsymbol{U}\boldsymbol{U}^* = \boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I},$$

对任意的 $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}^n$，

$$(\boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{U}^*\boldsymbol{U}\boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{y}),$$

即(60)成立，所以条件是必要的.

现假定(60)对任意的向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathscr{C}^n$ 都成立，其中，令 $\boldsymbol{y} = \boldsymbol{x}$ 得

$$(\boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{x}),$$

所以有

$$(\boldsymbol{U}^*\boldsymbol{U}\boldsymbol{x}, \boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{x}).$$

这就是说，对一切 $\boldsymbol{x} \in \mathscr{C}^n$，恒有

$$((\boldsymbol{U}^*\boldsymbol{U} - \boldsymbol{I})\boldsymbol{x}, \boldsymbol{x}) = 0. \tag{61}$$

算子 $\boldsymbol{U}^*\boldsymbol{U} - \boldsymbol{I}$ 显然为自伴算子，记之为 $\boldsymbol{A}$，

$$\boldsymbol{A} = \boldsymbol{U}^*\boldsymbol{U} - \boldsymbol{I}.$$

今证必有 $\boldsymbol{A} = \boldsymbol{0}$. 为此目的，我们可利用恒等式

$$(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{A}\boldsymbol{y}, \boldsymbol{x}) = (\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{y}), \boldsymbol{x}+\boldsymbol{y}) - (\boldsymbol{A}\boldsymbol{x}, \boldsymbol{x}) - (\boldsymbol{A}\boldsymbol{y}, \boldsymbol{y}). \tag{62}$$

因为 $\boldsymbol{A}$ 是自伴算子，所以 $(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}, \boldsymbol{A}\boldsymbol{y})$. 但是根据内积的性质，

$$(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}) = (\overline{\boldsymbol{y}, \boldsymbol{A}\boldsymbol{x}}) = (\overline{\boldsymbol{A}\boldsymbol{y}, \boldsymbol{x}}),$$

这就是说，$(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y})$ 和 $(\boldsymbol{A}\boldsymbol{y}, \boldsymbol{x})$ 为共轭复数. 于是，(62)的左端可以表示成

$$(\boldsymbol{Ax}, \boldsymbol{y})+(\boldsymbol{Ay}, \boldsymbol{x})=2\operatorname{Re}(\boldsymbol{Ax}, \boldsymbol{y}).$$

记号"Re"表示取实数部. 同时，由(61)可知，(62)的右端化为0，所以必有

$$\operatorname{Re}(\boldsymbol{Ax}, \boldsymbol{y})=0.$$

令
$$|(\boldsymbol{Ax}, \boldsymbol{y})|=\theta(\boldsymbol{Ax}, \boldsymbol{y})=(\boldsymbol{A}\theta\boldsymbol{x}, \boldsymbol{y}),$$

其中 θ 是某一个模为1的复数. 以 $\theta\boldsymbol{x}$ 代替 $\operatorname{Re}(\boldsymbol{Ax}, \boldsymbol{y})$ 中的 $\boldsymbol{x}$ 就得到关系

$$\begin{aligned}0=\operatorname{Re}(\boldsymbol{A}(\theta\boldsymbol{x}), \boldsymbol{y})&=\operatorname{Re}[\theta(\boldsymbol{Ax}, \boldsymbol{y})]\\&=\operatorname{Re}|(\boldsymbol{Ax}, \boldsymbol{y})|=|(\boldsymbol{Ax}, \boldsymbol{y})|.\end{aligned}$$

由 $\boldsymbol{x}$, $\boldsymbol{y}$ 的任意性可知

$$\boldsymbol{A}=\boldsymbol{0},$$

即 $\boldsymbol{U}^*\boldsymbol{U}=\boldsymbol{I}$.

同理可证 $\boldsymbol{UU}^*=\boldsymbol{I}$，即 $\boldsymbol{U}$ 是一个酉算子. **证毕.**

推论 条件(60)等价于：对任意的 $\boldsymbol{x}\in\mathscr{C}^n$，恒有

$$(\boldsymbol{Ux}, \boldsymbol{Ux})=(\boldsymbol{x}, \boldsymbol{x}).$$

这就是说，酉算子保持向量的长度不变.

酉算子的矩阵表示 在酉空间 $\mathscr{C}^n$ 中取定基底后，$\boldsymbol{U}$ 算子可以表示成一个 $n\times n$ 阶矩阵. 因为算子积的矩阵表示为算子矩阵表示之积，故有

$$U^*U=UU^*=I_n, \tag{63}$$

I_n 表示 $n\times n$ 单位阵. 从而可知酉算子的矩阵表示具有关系

$$U^*=U^{-1}.$$

§6 投影算子及其矩阵表示

6-1 投影算子的定义和基本性质

现在我们假定酉空间 $\mathscr{C}^n$ 已经按某种规则确定了一种直接和分解：

$$\mathscr{C}^n = \mathscr{W} \oplus \mathscr{V}. \tag{64}$$

于是，空间 $\mathscr{C}^n$ 中的每一个向量 $\boldsymbol{x}$ 都可以唯一地表示成

$$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{v}, \quad \boldsymbol{w} \in \mathscr{W}, \quad \boldsymbol{v} \in \mathscr{V} \tag{65}$$

的形式．今规定从 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 内的线性算子 $\boldsymbol{P}$ 如下：

$$\boldsymbol{P}\boldsymbol{x} = \boldsymbol{w}. \tag{66}$$

我们称 $\boldsymbol{P}$ 为空间 $\mathscr{C}^n$ 的直接和分解(64)所确定的一个**投影算子**．更详细一些，$\boldsymbol{P}$ 是从酉空间 $\mathscr{C}^n$ 沿子空间 $\mathscr{V}$ 到子空间 $\mathscr{W}$ 上的投影算子．显然，若 $\boldsymbol{x} \in \mathscr{W}$，则 $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}$，从而可知

$$\boldsymbol{P}\mathscr{W} \subset \mathscr{W}, \tag{67}$$

这就是说，$\mathscr{W}$ 是算子 $\boldsymbol{P}$ 的一个**不变子空间**．

我们称(66)中的 $\boldsymbol{w}$ 为向量 $\boldsymbol{x}$ 在子空间 $\mathscr{W}$ 中的像，或 $\boldsymbol{x}$ 的投影．而 $\mathscr{W}$ 实际上就是算子 $\boldsymbol{P}$ 的像空间 $\mathscr{R}(\boldsymbol{P})$．因为，对一切 $\boldsymbol{v} \in \mathscr{V}$，恒有

$$\boldsymbol{P}\boldsymbol{v} = \boldsymbol{0}, \tag{68}$$

所以，子空间 $\mathscr{V}$ 就是算子 $\boldsymbol{P}$ 的零空间 $\mathscr{N}(\boldsymbol{P})$．于是空间 $\mathscr{C}^n$ 的直接和分解(66)也可以表示成

$$\mathscr{C}^n = \mathscr{R}(\boldsymbol{P}) \oplus \mathscr{N}(\boldsymbol{P}). \tag{69}$$

容易证明，算子 $\boldsymbol{P}$ 是一个线性算子．

定理 12 从酉空间 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 内的线性算子 $\boldsymbol{P}$ 是投影算子的充分必要条件为：

$$\boldsymbol{P}^2 = \boldsymbol{P}, \tag{70}$$

即，$\boldsymbol{P}$ 是一个等幂算子．

证明 设 $\boldsymbol{P}$ 是空间分解(66)所确定的投影算子，$\boldsymbol{x} \in \mathscr{C}^n$，$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{v}$．于是

$$\boldsymbol{P}^2\boldsymbol{x} = \boldsymbol{P}(\boldsymbol{P}\boldsymbol{x}) = \boldsymbol{P}\boldsymbol{w} = \boldsymbol{w} = \boldsymbol{P}\boldsymbol{x},$$

所以 $\boldsymbol{P}$ 是一个等幂算子．

现在我们假定 $\boldsymbol{P}$ 是一个等幂算子．对空间 $\mathscr{C}^n$ 作直接和

分解

$$\mathscr{C}^n=\mathscr{R}(\boldsymbol{P})\oplus\mathscr{N}(\boldsymbol{P}).$$

设 $\boldsymbol{w}\in\mathscr{R}(\boldsymbol{P})$,则必存在 $\boldsymbol{x}\in\mathscr{C}^n$ 使

$$\boldsymbol{P}\boldsymbol{x}=\boldsymbol{w}.$$

据 $\boldsymbol{P}$ 的等幂性可知

$$\boldsymbol{P}\boldsymbol{w}=\boldsymbol{P}^2\boldsymbol{x}=\boldsymbol{P}\boldsymbol{x}=\boldsymbol{w}.$$

这就证明了 $\boldsymbol{P}$ 是一个投影算子. 证毕.

定义 15 空间分解(64)定义了从 $\mathscr{C}^n$ 沿 $\mathscr{V}$ 到 $\mathscr{W}$ 上的投影算子 $\boldsymbol{P}$,根据分解(64)的对称性,我们同时也可以定义从空间 $\mathscr{C}^n$ 沿子空间 $\mathscr{W}$ 到子空间 $\mathscr{V}$ 上的投影算子 $\boldsymbol{Q}$, 即, 对(65)中 $\boldsymbol{x}$ 的相应分解 $\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v}$,令

$$\boldsymbol{Q}\boldsymbol{x}=\boldsymbol{v}, \tag{71}$$

显然 $\boldsymbol{P}+\boldsymbol{Q}=\boldsymbol{I}$,这就是说,用投影算子 $\boldsymbol{P}$ 可以把 $\boldsymbol{Q}$ 表示出来

$$\boldsymbol{Q}=\boldsymbol{I}-\boldsymbol{P}. \tag{72}$$

我们称 $\boldsymbol{Q}$ 是 $\boldsymbol{P}$ 的补投影算子. 根据定义的对称性,$\boldsymbol{P}$ 本身也是 $\boldsymbol{Q}$ 的补投影算子.

投影的几何意义 今以三维欧氏空间 $\mathscr{R}^3$ 为例来加以说明. 设$\{\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3\}$是 $\mathscr{R}^3$ 的一个基底,令

$$\mathscr{W}=\mathrm{Span}\{\boldsymbol{g}_1, \boldsymbol{g}_2\}$$

$$\mathscr{V}=\mathrm{Span}\{\boldsymbol{g}_3\},$$

于是对 $\mathscr{R}^3$ 便可作直接和分解

$$\mathscr{R}^3=\mathscr{W}\oplus\mathscr{V},$$

向量 $\boldsymbol{x}$ 在子空间 $\mathscr{W}$ 上的投影为 $\boldsymbol{P}\boldsymbol{x}=\boldsymbol{w}$, 在子空间 $\mathscr{V}$ 上的投影为 $\boldsymbol{Q}\boldsymbol{x}=\boldsymbol{v}$,如图 1-1 所示.

6-2 直交投影算子 设 $\mathscr{W}$ 是 n 维酉空间 $\mathscr{C}^n$ 中的一个子空间, $\mathscr{W}^{\perp}$ 为其直交补空间. 我们有直交分解

$$\mathscr{C}^n=\mathscr{W}\oplus\mathscr{W}^{\perp} \tag{73}$$

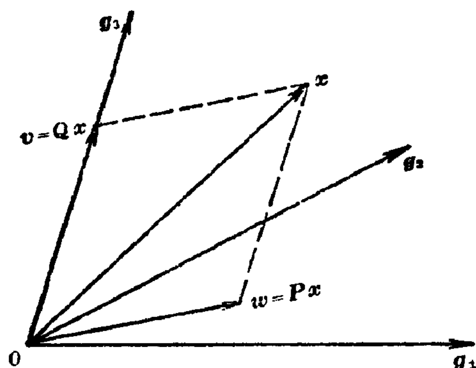


图 1-1

由子空间 $\mathscr{W}$ 所确定的这种分解，规定了从 $\mathscr{C}^n$ 沿 $\mathscr{W}^\perp$ 到 $\mathscr{W}$ 的投影算子，称之为直交投影算子．由于分解(73)为 $\mathscr{W}$ 所唯一确定，所以这种投影算子可以记成 $\boldsymbol{P}_{\mathscr{W}}$，以示它与 $\mathscr{W}$ 的关系．$\boldsymbol{P}_{\mathscr{W}}$ 的直交补投影算子可以表示为 $\boldsymbol{P}_{\mathscr{W}^\perp}$．

从酉空间 $\mathscr{C}^n$ 到子空间 $\mathscr{W}$ 上的投影算子不是唯一的，而 $\boldsymbol{P}_{\mathscr{W}}$ 则是唯一的．

如果把空间 $\mathscr{C}^n$ 进一步分解成 k 个互相直交子空间 $\mathscr{W}_1, \cdots, \mathscr{W}_k$ 的直接和

$$\mathscr{C}^n=\mathscr{W}_1 \oplus \mathscr{W}_2 \oplus \cdots \oplus \mathscr{W}_k, \tag{74}$$

这种分解就定义了 k 个直交投影算子 $\boldsymbol{P}_{\mathscr{W}_i}$, $i=1, 2, \cdots, k$. 我们有关系

$$\boldsymbol{P}_{\mathscr{W}_1}+\boldsymbol{P}_{\mathscr{W}_2}+\cdots+\boldsymbol{P}_{\mathscr{W}_k}=\boldsymbol{I}. \tag{75}$$

前面已经讲过，投影算子必为等幂算子，反之亦然．直交投影算子当然也必须具有等幂性，因为它是投影算子．但仅有等幂性还不够，它需要具有更进一步的性质．

定理 13 定义在酉空间 $\mathscr{C}^n$ 上的等幂算子 $\boldsymbol{P}$ 为直交投影算子的充分必要条件为：$\boldsymbol{P}$ 是自伴算子．

证明 今先证明直交投影算子 $\boldsymbol{P}$ 必为自伴算子.

和算子 $\boldsymbol{P}$ 相应的空间分解为

$$\mathscr{C}^n=\mathscr{R}(\boldsymbol{P})\oplus\mathscr{R}^{\perp}(\boldsymbol{P})=\mathscr{R}(\boldsymbol{P})\oplus\mathscr{N}(\boldsymbol{P}). \tag{76}$$

设 $\boldsymbol{x},\boldsymbol{y}$ 为 $\mathscr{C}^n$ 中任意两个向量,且

$$\boldsymbol{x}=\boldsymbol{w}_1+\boldsymbol{v}_1,\quad \boldsymbol{y}=\boldsymbol{w}_2+\boldsymbol{v}_2,$$

$$\boldsymbol{w}_1,\ \boldsymbol{w}_2\in\mathscr{R}(\boldsymbol{P}),\ \boldsymbol{v}_1,\ \boldsymbol{v}_2\in\mathscr{N}(\boldsymbol{P}).$$

于是 $$(\boldsymbol{Px},\ \boldsymbol{y})=(\boldsymbol{Pw}_1,\ \boldsymbol{w}_2+\boldsymbol{v}_2)=(\boldsymbol{w}_1,\ \boldsymbol{w}_2),$$

$$(\boldsymbol{x},\ \boldsymbol{Py})=(\boldsymbol{w}_1+\boldsymbol{v}_1,\ \boldsymbol{Pw}_2)=(\boldsymbol{w}_1,\ \boldsymbol{w}_2),$$

即,等式 $$(\boldsymbol{Px},\ \boldsymbol{y})=(\boldsymbol{x},\ \boldsymbol{Py})$$

恒成立,所以 $\boldsymbol{P}$ 必为自伴算子.

反之,若假定等幂算子 $\boldsymbol{P}$ 为自伴算子,则可以证明它必为一个直交投影算子. 就是说,对任意的 $\boldsymbol{v}\in\mathscr{R}^{\perp}(\boldsymbol{P})$,必有 $\boldsymbol{Pv}=0$.

对任意的 $\boldsymbol{x}\in\mathscr{C}^n$,根据 $\boldsymbol{P}$ 的自伴性,恒有

$$0=(\boldsymbol{Px},\ \boldsymbol{v})=(\boldsymbol{x},\ \boldsymbol{Pv}),$$

即,等式 $$(\boldsymbol{x},\ \boldsymbol{Pv})=0$$

对一切 $\boldsymbol{x}$ 都成立,所以必须 $\boldsymbol{Pv}=0$,这就证明 $\boldsymbol{P}$ 是一个直交投影算子. 证毕.

6-3 投影算子的矩阵表示

现在我们来考虑在空间中建立坐标系以后,如何把给定的投影算子用相应的矩阵表示出来的问题,首先我们考虑一般的投影算子. 设 $\boldsymbol{P}$ 是空间的直接和分解

$$\mathscr{C}^n=\mathscr{W}\oplus\mathscr{V}$$

所确定的从空间 $\mathscr{C}^n$ 沿子空间 $\mathscr{V}$ 到子空间 $\mathscr{W}$ 上的投影算子,并假定

$$\dim\mathscr{W}=r,\quad \dim\mathscr{V}=n-r.$$

在子空间 $\mathscr{W}$ 和 $\mathscr{V}$ 中分别取定基底

$$\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_r \text{ 和 } \boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n-r}.$$

这两组基底联合起来便构成空间 $\mathscr{C}^n$ 的一个基底．此外，我们还假定在空间 $\mathscr{C}^n$ 中已经建立了一个坐标系，而记号 $\boldsymbol{w}_i, \boldsymbol{v}_j$ 同时也是它们自己在这个坐标系中的表示．记 P 为投影算子 $\boldsymbol{P}$ 在给定坐标系中的矩阵表示，它是一个 $n\times n$ 阶矩阵．根据投影算子的定义，应有

$$\begin{aligned} \boldsymbol{P}\boldsymbol{w}_i &= \boldsymbol{w}_i, \quad i=1, 2, \cdots, r, \\ \boldsymbol{P}\boldsymbol{v}_j &= 0, \quad j=1, 2, \cdots, n-r. \end{aligned} \tag{77}$$

依次用 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_r$ 和 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n-r}$ 为列向量作 $n\times r$ 阶及 $n\times(n-r)$ 阶矩阵

$$W=[\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_r], \quad V=[\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{n-r}], \tag{78}$$

令

$$C=[W \,\vdots\, V] \tag{79}$$

于是有

$$\begin{aligned} PC &= [P\boldsymbol{w}_1, \cdots, P\boldsymbol{w}_r, P\boldsymbol{v}_1, \cdots, P\boldsymbol{v}_{n-r}] \\ &= [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_r, 0, \cdots, 0] = [W \,\vdots\, 0]. \end{aligned} \tag{80}$$

因为 C 的 n 个列向量线性独立，所以矩阵 C 非奇异，即逆矩阵 C^{-1} 存在．用它右乘(80)的两端，便得到投影算子 P 的矩阵表示

$$P=[W \,|\, 0]C^{-1}=[W \,\vdots\, 0][W \,|\, V]^{-1}. \tag{81}$$

于特例，如果用 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_r, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-r}$ 作为基底建立空间 $\mathscr{C}^n$ 的坐标系，那么，在这个坐标系中，$\boldsymbol{w}_i, \boldsymbol{v}_j$ 的表示为

$$\boldsymbol{w}_1=\begin{bmatrix}1\\0\\ \vdots\\ \vdots\\ \vdots\\ 0\end{bmatrix}, \cdots, \boldsymbol{w}_r=\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}, \quad \boldsymbol{v}_1=\begin{bmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{bmatrix}, \cdots, \quad \boldsymbol{v}_{n-r}=\begin{bmatrix}0\\ \vdots\\ \vdots\\ \vdots\\ 0\\ 1\end{bmatrix},$$

从而 $$W=\begin{bmatrix} I_r \\ 0 \end{bmatrix},\quad V=\begin{bmatrix} 0 \\ I_{n-r} \end{bmatrix},$$
其中的 I_r 和 I_{n-r} 分别表示 $r\times r$ 阶和 $(n-r)\times(n-r)$ 阶单位阵。这时
$$[W\mid V]=I_n,\quad [W\mid V]^{-1}=I_n,$$
从而投影算子的矩阵表示便简化为
$$P=\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}. \tag{82}$$

现在我们来考虑直交投影算子 $P_{\mathscr{W}}$ 的矩阵表示。为简单起见，我们在 $\mathscr{C}^n$ 中取一组直交基底建立坐标系，而 w_i、v_j 同时也表示它们在这个直角坐标系中相应的坐标向量，因为在直交投影的情形
$$\mathscr{V}=\mathscr{W}^{\perp},$$
所以 $\quad w_i\perp v_j,\ i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ n-r.$

于是有 $$W^*V=0,$$
右端是 $r\times(n-r)$ 阶零矩阵。在这个坐标系中，直交投影算子 $P_{\mathscr{W}}$ 的矩阵表示可推导如下。

因为
$$([W\mid V]^*[W\mid V])^{-1}=[W\mid V]^{-1}([W\mid V]^*)^{-1}$$
$$=\left(\begin{bmatrix} W^* \\ V^* \end{bmatrix}[W\mid V]\right)^{-1}=\begin{bmatrix} W^*W & 0 \\ 0 & V^*V \end{bmatrix}^{-1}$$
$$=\begin{bmatrix} (W^*W)^{-1} & 0 \\ 0 & (V^*V)^{-1} \end{bmatrix},$$
所以
$$[W\mid V]^{-1}=\begin{bmatrix} (W^*W)^{-1} & 0 \\ 0 & (V^*V)^{-1} \end{bmatrix}[W\mid V]^*$$
$$=\begin{bmatrix} (W^*W)^{-1}W^* \\ (V^*V)^{-1}V^* \end{bmatrix},$$

从而便得到 $\boldsymbol{P}_{\mathscr{W}}$ 的矩阵表示为

$$\begin{aligned}\boldsymbol{P}_{\mathscr{W}} &= [W|0]\,[W|V]^{-1} \\ &= [W|0]\begin{bmatrix}(W^*W)^{-1} & W^* \\ (V^*V)^{-1} & V^*\end{bmatrix} \\ &= W(W^*W)^{-1}W^*.\end{aligned} \tag{83}$$

如果 $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_r, \boldsymbol{v}_1, \cdots, \boldsymbol{v}_{n-r}$ 是一个标准直交系，则

$$W^*W = I_r,\quad V^*V = I_{n-r},$$

这时投影算子 $\boldsymbol{P}_{\mathscr{W}}$ 的矩阵表示就简化为

$$P_{\mathscr{W}} = WW^*. \tag{84}$$

6-4 投影算子的秩和直交投影算子的单秩分解

对于由空间分解 (64) 所确定的一般的投影算子 $\boldsymbol{P}$，其秩为

$$\operatorname{rank}(\boldsymbol{P}) = \dim R(\boldsymbol{P}) = \dim \mathscr{W} = r.$$

对于任何非零向量 $\boldsymbol{w} \in \mathscr{W}$，我们有

$$\boldsymbol{P}\boldsymbol{w} = \boldsymbol{w},$$

这就是说，子空间 $\mathscr{W}$ 中的任何非零向量 $\boldsymbol{w}$ 都是算子 $\boldsymbol{P}$ 的特征向量，它们相应的特征值都是 1. 而对 $\mathscr{V}$ 中的任何非零向量 $\boldsymbol{v}$，恒有

$$\boldsymbol{P}\boldsymbol{v} = \boldsymbol{0}.$$

这就是说，$\mathscr{V}$ 中的非零向量都是算子 $\boldsymbol{P}$ 的与零特征值相应的特征向量. 因为

$$\dim \mathscr{W} = r,\quad \dim \mathscr{V} = n - r,$$

所以，对应于特征值 1 只能有 r 个线性独立的特征向量，而对应于特征值零则只能有 $n-r$ 个线性独立的特征向量. 现在，我们对直交投影算子 $\boldsymbol{P}_{\mathscr{W}}$ 再作进一步的研究，并给出它的谱分解公式.

设 $\dim \mathscr{W} = r$. 在子空间 $\mathscr{W}$ 中取出一个直交系

$$\boldsymbol{w}_1,\ \boldsymbol{w}_2,\ \cdots,\ \boldsymbol{w}_r$$

作为 $\mathscr{W}$ 的基底，并记由向量 $\boldsymbol{w}_i$，$i=1, 2, \cdots, r$，所张成的一维子空间为 $\mathscr{W}_i$，即，

$$\mathscr{W}_i=\mathrm{Span}\{\boldsymbol{w}_i\},\ i=1, 2, \cdots, r, \tag{85}$$

由子空间 $\mathscr{W}_i$ 所确定的直交投影算子记之为 $\boldsymbol{P}_i$. 于是

$$\left.\begin{aligned} \boldsymbol{P}_i\boldsymbol{w}_i&=\boldsymbol{w}_i, \\ \boldsymbol{P}_i\boldsymbol{w}_j&=\boldsymbol{0},\ j\neq i \end{aligned}\right\}\quad i, j=1, 2, \cdots, r, \tag{86}$$

并且，对任意的 $\boldsymbol{x}\in\mathscr{W}^{\perp}$，恒有

$$\boldsymbol{P}_i\boldsymbol{x}=0,\ i=1, 2, \cdots, r. \tag{87}$$

根据(85)和(86)显然可知，由 r 维子空间 $\mathscr{W}$ 所确定的直交投影算子 $\boldsymbol{P}_{\mathscr{W}}$ 可以表示成 r 个直交投影算子 $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_r$ 的和

$$\boldsymbol{P}_{\mathscr{W}}=\boldsymbol{P}_1+\boldsymbol{P}_2+\cdots+\boldsymbol{P}_r, \tag{88}$$

这就是直交投影算子 $\boldsymbol{P}_{\mathscr{W}}$ 的**单秩分解公式**.

如果我们在空间 $\mathscr{C}^n$ 中选定一个标准直交系作为基底以建立空间 $\mathscr{C}^n$ 的坐标系，并假定 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_r$ 是前述 $\mathscr{W}$ 的一个标准直交基底，而记号 $\boldsymbol{w}_i$ 本身同时也表示关于 $\mathscr{C}^n$ 中所选标准直交基底下的坐标向量，则直交投影算子 $\boldsymbol{P}_i$，$i=1, 2, \cdots, r$ 的矩阵表示为

$$P_i=\boldsymbol{w}_i\boldsymbol{w}_i^*,\quad i=1, 2, \cdots, r, \tag{89}$$

从而直交投影算子 $\boldsymbol{P}_{\mathscr{W}}$ 的矩阵表示 $P_{\mathscr{W}}$ 便可以表示为

$$P_{\mathscr{W}}=\boldsymbol{w}_1\boldsymbol{w}_1^*+\boldsymbol{w}_2\boldsymbol{w}_2^*+\cdots+\boldsymbol{w}_r\boldsymbol{w}_r^*, \tag{90}$$

因为

$$P_i\boldsymbol{w}_i=\boldsymbol{w}_i,\ i=1, 2, \cdots, r,$$

$$P_i\boldsymbol{x}=\boldsymbol{0},\ 若\ \boldsymbol{x}\in\mathscr{W}_i^{\perp},$$

可知 $\boldsymbol{w}_i$ 是 P_i 的特征向量，相应的特征值为1，由于其秩为1，所以其余 $n-1$ 个特征向量都对应于特征值零. 由此还可以进一步看出，$P_{\mathscr{W}}$ 有 r 个特征向量 $\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_r$ 对应于特征值1，而其余 $n-r$ 个特征向量则对应于特征值零。因此，我

们称(90)为 $P_{\mathscr{V}}$ 的谱分解.

§7 向量和算子的范数

7-1 向量的范数及其基本性质

设 $\mathscr{V}$ 是定义在复数域 $\mathscr{C}$ (或实数域 $\mathscr{R}$) 上的 n 维酉空间, 在 $\mathscr{V}$ 中定义了向量的内积. 前面已经讲过, 利用内积可定义向量 $\boldsymbol{x}$ 的范数

$$\|\boldsymbol{x}\|^2=(\boldsymbol{x},\ \boldsymbol{x}). \tag{91}$$

现在我们进一步来研究向量范数的基本属性.

定理 14 由(91)定义的向量范数具有下列性质:

1) $\|\boldsymbol{x}\|\geqslant 0$, 只有 $\boldsymbol{x}=\boldsymbol{0}$ 时才有 $\|\boldsymbol{x}\|=0$;

2) $\|\alpha\boldsymbol{x}\|=|\alpha|\,\|\boldsymbol{x}\|$, 对任意的 $\alpha\in\mathscr{C}$, $\boldsymbol{x}\in\mathscr{V}$;

3) 对任意的向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$, 恒有

$$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|.$$

证明 性质 1)～2) 是明显的, 现在只需要证明 3) 就行了, 为此, 我们先证明向量内积满足 Schwarz 不等式:

$$|(\boldsymbol{x},\ \boldsymbol{y})|\leqslant\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|. \tag{92}$$

因为根据内积的性质, 对任意的向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 以及复数 α, 恒有

$$(\boldsymbol{x}+\alpha\boldsymbol{y},\ \boldsymbol{x}+\alpha\boldsymbol{y})\geqslant 0,$$

即 $$(\boldsymbol{x},\ \boldsymbol{x})+\bar{\alpha}(\boldsymbol{x},\ \boldsymbol{y})+\alpha(\boldsymbol{y},\ \boldsymbol{x})+|\alpha|^2(\boldsymbol{y},\ \boldsymbol{y})\geqslant 0.$$

若 $\boldsymbol{y}=\boldsymbol{0}$, 则 Schwarz 不等式显然成立, 今假定 $\boldsymbol{y}\neq\boldsymbol{0}$, 这时可令

$$\alpha=-\frac{(\boldsymbol{x},\ \boldsymbol{y})}{(\boldsymbol{y},\ \boldsymbol{y})},$$

把 α 代入上面的不等式中, 得

$$(\boldsymbol{x},\ \boldsymbol{x})-\frac{|(\boldsymbol{x},\ \boldsymbol{y})|^2}{(\boldsymbol{y},\ \boldsymbol{y})}\geqslant 0,$$

即 $$|(\boldsymbol{x},\ \boldsymbol{y})|^2\leqslant(\boldsymbol{x},\ \boldsymbol{x})(\boldsymbol{y},\ \boldsymbol{y})=\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2,$$

或 $$|(\boldsymbol{x},\ \boldsymbol{y})|\leqslant\|\boldsymbol{x}\|\|\boldsymbol{y}\|,$$

从而便证明了 Schwarz 不等式.

现在来证明范数的性质 3). 因为

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2&=(\boldsymbol{x}+\boldsymbol{y},\ \boldsymbol{x}+\boldsymbol{y})=(\boldsymbol{x},\ \boldsymbol{x})+(\boldsymbol{y},\ \boldsymbol{x})\\&\quad+(\boldsymbol{x},\ \boldsymbol{y})+(\boldsymbol{y},\ \boldsymbol{y})\leqslant\|\boldsymbol{x}\|^2+2\|\boldsymbol{x}\|\|\boldsymbol{y}\|+\|\boldsymbol{y}\|^2\\&=(\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)^2,\end{aligned}$$

所以必有 $$\|\boldsymbol{x}+\boldsymbol{y}\|\leqslant\|\boldsymbol{x}\|+\|\boldsymbol{y}\|.$$ 证毕.

以上我们证明了由内积所定义的向量范数所具有的一些基本性质. 一般来说,定义向量范数并不一定要借助于内积,可定义$\|\boldsymbol{x}\|$为向量 $\boldsymbol{x}$ 的一个实值函数, 使它具有前述定理中的性质 1)～3).

在 n 维向量空间 $\mathscr{V}$ 中建立了一个标准直交基底后,任何向量 $\boldsymbol{x}$ 都可以表示成相应的坐标向量

$$\boldsymbol{x}=\begin{bmatrix}\xi_1\\ \vdots\\ \xi_n\end{bmatrix}.$$

前述向量范数 $\|\boldsymbol{x}\|$ 便可以表示成

$$\|\boldsymbol{x}\|=(|\xi_1|^2+\cdots+|\xi_n|^2)^{\frac{1}{2}}. \tag{93}$$

实际上,一般也可以规定

$$(|\xi_1|^p+\cdots+|\xi_n|^p)^{\frac{1}{p}}$$

来作为范数的定义,可以证明,只要 $p\geqslant 1$,范数的性质 1)～3) 都成立. 这种范数常记作 $\|\boldsymbol{x}\|_p$. 例如,当 $p=1$ 时

$$\|\boldsymbol{x}\|_1=|\xi_1|+\cdots+|\xi_n|. \tag{94}$$

当 $p=\infty$ 时,可证

$$\|x\|_\infty=\max_{1\leqslant i\leqslant n}|\xi_i|. \tag{95}$$

而前面用内积所定义的范数(93)实际上就是这种 p-范数的一种特例，即 $\|x\|_2$. 因为本书主要讨论广义逆矩阵和相应的最小二乘问题，对我们来说重要的是 $\|x\|_2$，所以，除非特别声明，记号 $\|x\|$ 总表示向量的欧氏范数——即 2-范数.

7-2 线性算子的范数

有了向量范数，我们就可以规定相应的算子范数．为此我们先证明

定理 15 设 $\mathscr{V}_1$ 和 $\mathscr{V}_2$ 是定义在复数(或实数)域上的有限维酉空间，A 是从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子，则 A 必有界，即对一切 $x\in\mathscr{V}_1$，恒存在正常数 k，使

$$\|Ax\|\leqslant k\|x\|. \tag{96}$$

证明 设 $\dim\mathscr{V}_1=n$，取 $\mathscr{V}_1$ 中的一个标准直交基底 $\{x_1, \cdots, x_n\}$，则任何向量 $x\in\mathscr{V}_1$ 都可以表示成

$$x=\sum_{i=1}^{n}(x, x_i)x_i.$$

因为 $\|x_i\|=1$，$i=1, 2, \cdots, n$，于是便得到

$$\begin{aligned}\|Ax\|&=\left\|A\sum_{i=1}^{n}(x, x_i)x_i\right\|=\left\|\sum_{i=1}^{n}(x, x_i)Ax_i\right\|\\&\leqslant\sum_{i=1}^{n}|(x, x_i)|\,\|Ax_i\|\leqslant\sum_{i=1}^{n}\|x\|\cdot\|x_i\|\,\|Ax_i\|\\&\leqslant\left(\sum_{i=1}^{n}\|Ax_i\|\right)\|x\|.\end{aligned}$$

令
$$k=\sum_{i=1}^{n}\|Ax_i\|,$$
便得到
$$\|Ax\|\leqslant k\|x\|.$$

从而便证明了线性算子 A 的有界性．证毕．

根据算子的有界性，立刻便可以证明线性算子对向量 x

的连续性，即，当 $\|\delta \boldsymbol{x}\|\to 0$ 时，

$$\|\boldsymbol{A}(\boldsymbol{x}+\delta \boldsymbol{x})\|\to\|\boldsymbol{A}\boldsymbol{x}\|.$$

因为 $\|\boldsymbol{A}(\boldsymbol{x}+\delta\boldsymbol{x})-\boldsymbol{A}\boldsymbol{x}\|=\|\boldsymbol{A}\delta\boldsymbol{x}\|\leqslant k\|\delta\boldsymbol{x}\|$，

而 $\big|\|\boldsymbol{A}(\boldsymbol{x}+\delta\boldsymbol{x})\|-\|\boldsymbol{A}\boldsymbol{x}\|\big|\leqslant\|\boldsymbol{A}(\boldsymbol{x}+\delta\boldsymbol{x})-\boldsymbol{A}\boldsymbol{x}\|$.

定义 16

$$\|\boldsymbol{A}\|=\sup_{\boldsymbol{x}\neq 0}\frac{\|\boldsymbol{A}\boldsymbol{x}\|}{\|\boldsymbol{x}\|}. \tag{97}$$

因为以 $\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$ 代 $\boldsymbol{x}$，上式右端的分式不变，所以(97)可以改写为

$$\|\boldsymbol{A}\|=\sup_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|.$$

另外，因为 $\boldsymbol{A}\boldsymbol{x}$ 是连续的，而单位球面

$$\|\boldsymbol{x}\|=1$$

是一个有界闭集，所以总存在 $\|\tilde{\boldsymbol{x}}\|=1$，使

$$\|\boldsymbol{A}\tilde{\boldsymbol{x}}\|=\sup_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|,$$

所以上式又可以改写为

$$\|\boldsymbol{A}\|=\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|. \tag{98}$$

这样定义的算子范数，显然具有和向量范数同样的性质：

1° $\|\boldsymbol{A}\|\geqslant 0$，只有 $\boldsymbol{A}=\boldsymbol{0}$ 时才有 $\|\boldsymbol{0}\|=0$；

2° 对任意复数 α，$\|\alpha\boldsymbol{A}\|=|\alpha|\,\|\boldsymbol{A}\|$；

3° 设 $\boldsymbol{B}$ 也是从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子，则

$$\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|.$$

此外，还不难证明下列结论：

4° 设 $\boldsymbol{I}$ 为从 $\mathscr{V}_1$ 到 $\mathscr{V}_1$ 的不变算子，即对任意 $\boldsymbol{x}\in\mathscr{V}_1$，恒有 $\boldsymbol{I}\boldsymbol{x}=\boldsymbol{x}$，则

$$\|\boldsymbol{I}\|=1;$$

5° 若 $\boldsymbol{P}$ 为定义在 $\mathscr{V}_1$ 上的投影算子，则

$$\|P\|=1;$$

6° 若 U 为从 $\mathscr{V}_1$ 到 $\mathscr{V}_1$ 的酉算子，则恒有

$$\|U\|=\|U^*\|=1.$$

7-3 算子范数的几个等价定义

前面所定义的算子范数，还可以表示成其它等价的形式。我们有

定理16 设 A 为从酉空间 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 的线性算子，则下列结论成立：

(i) $$\|A\|=\sup_{\substack{x\neq 0\\ y\neq 0}}\left\{\frac{|(Ax,\ y)|}{\|x\|\|y\|},\ x\in\mathscr{V}_1,\ y\in\mathscr{V}_2\right\};$$

(ii) $$\|A\|=\sup_{\|x\|=\|y\|=1}\{|(Ax,\ y)|,\ x\in\mathscr{V}_1,\ y\in\mathscr{V}_2\}.$$

证明 (i)、(ii)右端的表达式显然是等价的，今只需证明(i)与(ii)之一即可，记

$$r=\sup_{\substack{x\neq 0\\ y\neq 0}}\left\{\frac{|(Ax,\ y)|}{\|x\|\|y\|},\ x\in\mathscr{V}_1,\ y\in\mathscr{V}_2\right\},$$

因为 $|(Ax,\ y)|\leqslant\|Ax\|\|y\|\leqslant\|A\|\cdot\|x\|\cdot\|y\|$,

所以必有 $$r\leqslant\|A\|.$$

次设 x 为使 $Ax\neq 0$ 的任一向量(这里可假定 $A\neq 0$，否则定理显然)．于是 $x\neq 0$，令 $y=Ax$，则有

$$\frac{\|Ax\|}{\|x\|}=\frac{|(Ax,\ y)|}{\|x\|\|y\|}.$$

所以

$$\|A\|=\sup_{x\neq 0}\frac{\|Ax\|}{\|x\|}\leqslant\sup_{\substack{x\neq 0\\ y\neq 0}}\left\{\frac{|(Ax,\ y)|}{\|x\|\|y\|},\ x\in\mathscr{V}_1,\ y\in\mathscr{V}_2\right\},$$

从而可知 $$\|A\|\leqslant r,$$

于是就证明了等式 $\|A\|=r$．证毕．

根据这个定理，可以证明算子范数的另一重要性质，

7° 设 A^* 为 A 的伴随算子，则

$$\|A^*\| = \|A\|.$$

证明 任取二向量 $x \in \mathscr{V}_1$, $y \in \mathscr{V}_2$, $x \neq 0$, $y \neq 0$, 则

$$|(Ax, y)| = |(x, A^*y)| \leqslant \|A^*\| \|x\| \|y\|,$$

故有

$$\frac{|(Ax, y)|}{\|x\| \|y\|} \leqslant \|A^*\|,$$

这就表示

$$\|A\| \leqslant \|A^*\|.$$

交换 A 和 A^* 的位置，便得到相反的不等式

$$\|A^*\| \leqslant \|A\|.$$

所以必有 $\|A^*\| = \|A\|$. 证毕.

7-4 算子乘积的范数

设 $\mathscr{V}_1$, $\mathscr{V}_2$, $\mathscr{V}_3$ 都是有限维酉空间，A 为从 $\mathscr{V}_1$ 到 $\mathscr{V}_2$ 内的线性算子，B 为一个从 $\mathscr{V}_2$ 到 $\mathscr{V}_3$ 内的线性算子. 设 $x \in \mathscr{V}_1$, 则

$$y = Ax \in \mathscr{V}_2,$$

$$z = By \in \mathscr{V}_3,$$

今规定算子 B 和 A 的乘积为算子 BA,

$$(BA)x = By = z.$$

显然 BA 也是从 $\mathscr{V}_1$ 到 $\mathscr{V}_3$ 内的线性算子. 根据算子范数的意义，不难证明

8° $\|BA\| \leqslant \|A\| \|B\|.$

第2章

广义逆矩阵的算子理论

§1 线性算子方程的求解问题

设$\mathscr{C}^n$和$\mathscr{C}^m$为给定的n维和m维酉空间，$\boldsymbol{A}$为从$\mathscr{C}^n$到$\mathscr{C}^m$内的线性算子．今考虑线性算子方程

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}, \tag{1}$$

其中$\boldsymbol{b}\in\mathscr{C}^m$为给定的向量，而$\boldsymbol{x}\in\mathscr{C}^n$为待定向量．

定义1 若存在向量$\boldsymbol{x}\in\mathscr{C}^n$满足方程(1)，便称线性算子方程(1)为**相容**，否则就称(1)为**不相容**，或说它是一个**矛盾方程**．

关于线性算子方程的求解问题，常见的有以下几种情形．

1° 相容方程的求解问题．

如果线性算子方程(1)为相容，其解可能有无限多个，但在解族$\{\boldsymbol{x}\}$中，具有极小范数$\|\boldsymbol{x}\|$的解，可以证明是唯一的．

定义2 相容方程的解族中，具有极小范数的解，称为线性算子方程的**极小范数解**．

于是便有

2° 求相容方程的极小范数解的问题．

如果(1)不相容，即它是一个矛盾方程．这时(1)不存在通常意义下的解，但是可以考虑求$\boldsymbol{x}\in\mathscr{C}^n$，使

$$\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|=\text{极小} \tag{2}$$

的问题．于是便有

3° 求矛盾方程的最小二乘解问题．

定义3 我们称(2)是一个最小二乘问题，其解称为线性算子方程(1)的最小二乘解.

可以证明，最小二乘解总存在，但也可能有无限多个. 在(1)的最小二乘解集合中，具有极小范数的解则总是唯一的.

定义4 具有极小范数的最小二乘解，称之为线性算子方程(1)的极小最小二乘解.

于是便有

4° 求线性算子方程(1)的极小最小二乘解问题.

以上我们提出了四类线性算子方程的求解问题，关于各类问题的解的存在性和唯一性问题，以及其它性质，我们将在以后依次予以讨论.

§2 相容方程求解问题和相应的广义逆算子 A^-

2-1 线性算子方程的相容性条件和解集的构造

如果方程(1)存在有解 y，则

$$b = Ay \in \mathscr{C}^m,$$

这就是说 $b \in \mathscr{R}(A)$. 反之，若 b 属于算子 A 的像空间 $\mathscr{R}(A)$，显然方程(1)必有解，从而为相容，因此我们有

定理1 线性算子方程(1)为相容的充分必要条件是

$$b \in \mathscr{R}(A).$$

为了弄清楚相容方程(1)解集的构造，我们先来考虑和(1)相联系的齐次方程

$$Ax = 0.$$

显然，这个齐次方程的解集便是线性算子 A 的零空间 $\mathscr{N}(A)$.

定理2 设 x_0 为相容线性算子(1)的某一个特解，则它

的所有解的全体构成线性流形

$$\boldsymbol{x}_0+\mathcal{N}(\boldsymbol{A}). \tag{3}$$

证明 在 $\boldsymbol{A}$ 的零空间 $\mathcal{N}(\boldsymbol{A})$ 中任取一个向量 $\boldsymbol{u}$, 显然有

$$\boldsymbol{A}(\boldsymbol{x}_0+\boldsymbol{u})=\boldsymbol{A}\boldsymbol{x}_0=\boldsymbol{b},$$

这就是说，向量 $\boldsymbol{x}_0+\boldsymbol{u}$ 也是相容方程(1)的一个解. 换句话说，线性流形(3)中的一切向量都是方程(1)的解.

现在我们来证明方程(1)的任何一个解 $\boldsymbol{x}$, 都属于线性流形(3). 因为 $\boldsymbol{x}_0$、$\boldsymbol{x}$ 都是方程(1)的解，所以

$$\boldsymbol{A}\boldsymbol{x}_0=\boldsymbol{b},$$

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}.$$

两端分别相减便得到

$$\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_0)=\boldsymbol{0},$$

这就证明，向量 $\boldsymbol{x}$ 和 $\boldsymbol{x}_0$ 之差 $\boldsymbol{x}-\boldsymbol{x}_0\in\mathcal{N}(\boldsymbol{A})$. 令

$$\boldsymbol{u}=\boldsymbol{x}-\boldsymbol{x}_0,$$

便得到 $$\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{u}\in\{\boldsymbol{x}_0+\mathcal{N}(\boldsymbol{A})\}.$$

从而定理得证.

如果 $\dim\mathcal{R}(\boldsymbol{A})=r$, 则 $\dim\mathcal{N}(\boldsymbol{A})=n-r$. 由此可知，若从 $\mathcal{N}(\boldsymbol{A})$ 中选定一个基底 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_{n-r}$, 则方程(1)的任何解 $\boldsymbol{x}$ 都可以唯一地表示成

$$\boldsymbol{x}=\boldsymbol{x}_0+\sum_{i=1}^{n-r}\alpha_i\boldsymbol{u}_i, \quad \alpha_i\in\mathscr{C}. \tag{4}$$

2-2 相容算子方程求解和相应的广义逆算子 A^-

对于一般的 $n\times n$ 阶线性方程组求解，如果系数矩阵 A 非奇异，则可以用逆矩阵 A^{-1} 把解表示出来. 为了把这种思想推广到本章所考虑的线性算子方程，我们先对空间 $\mathscr{C}^n$ 作适当的直接和分解. 当线性算子 $\boldsymbol{A}$ 给定以后，它的零空间

$\mathscr{N}(\boldsymbol{A})$也就确定了．今假定子空间 S 是 $\mathscr{N}(\boldsymbol{A})$的一个补空间，一般说来，它不一定是 $\mathscr{N}(\boldsymbol{A})$的直交补空间．这时便有相应的直接和分解

$$\mathscr{C}^n = S \oplus \mathscr{N}(\boldsymbol{A}). \tag{5}$$

因为 $\dim \mathscr{R}(\boldsymbol{A}) = r$, $\dim \mathscr{N}(\boldsymbol{A}) = n-r$, 所以

$$\dim S = r,$$

即，子空间 S 的维数和象空间 $\mathscr{R}(\boldsymbol{A})$ 的维数相同．所以子空间 S 和子空间 $\mathscr{R}(\boldsymbol{A})$ 是同构的．

定理 3 我们有以下两个基本关系：

1) $\boldsymbol{A}S = \mathscr{R}(\boldsymbol{A})$;

2) 映射 $\boldsymbol{A}$ 从 S 到 $\mathscr{R}(\boldsymbol{A})$是一对一的．

证明 1) 显然 $\boldsymbol{A}S \subset \mathscr{R}(\boldsymbol{A})$．今只需证明，对 $\mathscr{R}(\boldsymbol{A})$ 中任意的向量 $\boldsymbol{y}$，总存在向量 $\boldsymbol{x} \in S$ 使 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ 即可．

因为 $\boldsymbol{y} \in \mathscr{R}(\boldsymbol{A})$，所以存在 $\boldsymbol{z} \in \mathscr{C}^n$, 使 $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}$．但向量 $\boldsymbol{z}$ 可唯一地表示成

$$\boldsymbol{z} = \boldsymbol{u} + \boldsymbol{v},$$

其中 $\boldsymbol{u} \in S$, $\boldsymbol{v} \in \mathscr{N}(\boldsymbol{A})$．所以

$$\boldsymbol{A}\boldsymbol{z} = \boldsymbol{A}(\boldsymbol{u}+\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{u} = \boldsymbol{y}.$$

这就证明了 S 中存在有与 $\boldsymbol{y}$ 对应的向量 $\boldsymbol{u}$，而且和 $\boldsymbol{y}$ 对应的 $\boldsymbol{u}$ 还是唯一的．

2) 现在只要证明和 $\mathscr{R}(\boldsymbol{A})$ 中两个不同的向量 $\boldsymbol{y}_1$ 和 $\boldsymbol{y}_2$ 相应的 $\boldsymbol{u}_1$ 和 $\boldsymbol{u}_2$（它们都属于 S）也不相同就行了．

因为 $$\boldsymbol{A}\boldsymbol{u}_1 = \boldsymbol{y}_1, \quad \boldsymbol{A}\boldsymbol{u}_2 = \boldsymbol{y}_2,$$

所以 $$\boldsymbol{A}(\boldsymbol{u}_1 - \boldsymbol{u}_2) = \boldsymbol{y}_1 - \boldsymbol{y}_2 \neq \boldsymbol{0},$$

从而可知 $\boldsymbol{u}_1 \neq \boldsymbol{u}_2$．这就证明了映射 $\boldsymbol{A}$ 从 S 到 $\mathscr{R}(\boldsymbol{A})$ 是一对一的，即是双向单值的．证毕．

在这个基础上，我们就可以证明

定理4 在$\mathscr{R}(\boldsymbol{A})$上存在有唯一的$\boldsymbol{A}$的逆算子$\boldsymbol{G}$，它把$\mathscr{R}(\boldsymbol{A})$映射为$S$，而且$\boldsymbol{G}$是线性算子.

证明 对于每一个$\boldsymbol{y}\in\mathscr{R}(\boldsymbol{A})$都存在唯一的向量$\boldsymbol{x}\in S$使

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y},$$

今规定

$$\boldsymbol{G}\boldsymbol{y}=\boldsymbol{x}. \tag{6}$$

算子$\boldsymbol{G}$就是$\boldsymbol{A}$的逆算子，其定义域为$\mathscr{R}(\boldsymbol{A})$，值域为$S$. 当子空间$S$确定以后，算子$\boldsymbol{G}$是唯一的.

现在我们来证明$\boldsymbol{G}$是一个线性算子. 因为，若

$$\boldsymbol{G}\boldsymbol{y}_1=\boldsymbol{x}_1,\ \boldsymbol{G}\boldsymbol{y}_2=\boldsymbol{x}_2,\ \boldsymbol{y}_1,\ \boldsymbol{y}_2\in\mathscr{R}(\boldsymbol{A}),\ \boldsymbol{x}_1,\ \boldsymbol{x}_2\in S,$$

则对任何复数α_1、α_2，由于$\boldsymbol{A}\boldsymbol{x}_1=\boldsymbol{y}_1$，$\boldsymbol{A}\boldsymbol{x}_2=\boldsymbol{y}_2$，恒有

$$\boldsymbol{A}(\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2)=\alpha_1\boldsymbol{A}\boldsymbol{x}_1+\alpha_2\boldsymbol{A}\boldsymbol{x}_2=\alpha_1\boldsymbol{y}_1+\alpha_2\boldsymbol{y}_2.$$

所以 $\boldsymbol{G}(\alpha_1\boldsymbol{y}_1+\alpha_2\boldsymbol{y}_2)=\alpha_1\boldsymbol{x}_1+\alpha_2\boldsymbol{x}_2=\alpha_1\boldsymbol{G}\boldsymbol{y}_1+\alpha_2\boldsymbol{G}\boldsymbol{y}_2.$

因此，算子$\boldsymbol{G}$是线性的. 证毕.

定义5 前面定理中的算子$\boldsymbol{G}$，称为算子$\boldsymbol{A}$的一种广义逆算子，记作$\boldsymbol{A}^-$，即

$$\boldsymbol{G}=\boldsymbol{A}^-. \tag{6'}$$

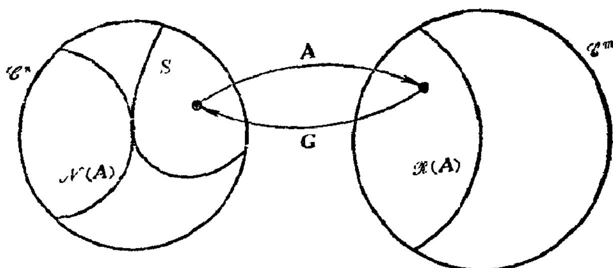


图2-1 确定广义逆算子$\boldsymbol{G}=\boldsymbol{A}^-$的示意图

这里还要着重指出，前述定理说$\boldsymbol{G}$是唯一的，但这是指当子空间S给定以后. 由于构成空间$\mathscr{C}^n$直接和分解的子空

间 S 并不是唯一的，故不同的 S，相应的广义逆算子 $\boldsymbol{G}$ 也不相同. 因此，对线性算子 $\boldsymbol{A}$ 来说，广义逆算子 $\boldsymbol{A}^-$ 不是唯一的.

有了广义逆算子 $\boldsymbol{G}$，相容方程(1)的解便可以表示为

$$\boldsymbol{x}=\boldsymbol{G}\boldsymbol{b},$$

或

$$\boldsymbol{x}=\boldsymbol{A}^-\boldsymbol{b}. \tag{7}$$

附注 若线性算子 $\boldsymbol{A}$ 为满秩，即 $\operatorname{rank}(\boldsymbol{A})=n$，这时，$\dim\mathcal{N}(\boldsymbol{A})=0$，而零空间 $\mathcal{N}(\boldsymbol{A})$ 中除零向量外无任何其它向量. 同时，$\dim S=n$，于是

$$S=\mathscr{C}^n.$$

这时，S 只能有一种选择. 对于一切 $\boldsymbol{x}\in\mathscr{C}^n$，恒有

$$\boldsymbol{G}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x},$$

所以

$$\boldsymbol{G}\boldsymbol{A}=\boldsymbol{I}, \tag{8}$$

$\boldsymbol{I}$ 为从 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 的不变算子. 因此，可以称 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的左逆.

2-3 广义逆算子 $\boldsymbol{A}^-$ 的特征性质——$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}=\boldsymbol{A}$

现设 $\boldsymbol{G}$ 为空间 $\mathscr{C}^n$ 的某一种直接和分解(5)所确定的线性算子 $\boldsymbol{A}$ 的一个广义逆算子 $\boldsymbol{A}^-$. 任取一向量 $\boldsymbol{x}\in\mathscr{C}^n$，令

$$\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v},\ \boldsymbol{u}\in S,\ \boldsymbol{v}\in\mathcal{N}(\boldsymbol{A})$$

及

$$\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}\boldsymbol{u}.$$

于是便有

$$\boldsymbol{G}\boldsymbol{b}=\boldsymbol{u}.$$

从而便有

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}\boldsymbol{G}\boldsymbol{b}=\boldsymbol{A}\boldsymbol{u}=\boldsymbol{A}(\boldsymbol{u}+\boldsymbol{v})=\boldsymbol{A}\boldsymbol{x},$$

于是，对一切 $\boldsymbol{x}\in\mathscr{C}^n$ 恒有

$$(\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}.$$

因此便得到关系

$$\boldsymbol{A}\boldsymbol{G}\boldsymbol{A}=\boldsymbol{A}. \tag{9}$$

反之，若 $\boldsymbol{G}$ 为定义在 $\mathscr{R}(\boldsymbol{A})$ 上的一个线性算子，其值域

在 $\mathscr{C}^n$ 内，即

$$\mathscr{R}(\boldsymbol{G})\subset\mathscr{C}^n.$$

如果算子 $\boldsymbol{G}$ 满足关系(9)，则 $\boldsymbol{G}$ 必是线性算子 $\boldsymbol{A}$ 的一个广义逆算子 $\boldsymbol{A}^-$. 今证明于下.

设 $\boldsymbol{b}\in\mathscr{R}(\boldsymbol{A})$, $\boldsymbol{Ax}=\boldsymbol{b}$. 于是

$$\boldsymbol{AGAx}=\boldsymbol{AGb}.$$

但是，
$$\boldsymbol{GAx}=\boldsymbol{Gb},$$

所以
$$\boldsymbol{AGAx}=\boldsymbol{AGb}=\boldsymbol{Ax}=\boldsymbol{b},$$

这就是说
$$\boldsymbol{A}(\boldsymbol{Gb})=\boldsymbol{b}.$$

所以，$\boldsymbol{Gb}$ 便是相容方程(1)的一个解. 从而可知 $\boldsymbol{G}$ 是 $\boldsymbol{A}$ 的一个广义逆算子 $\boldsymbol{A}^-$.

定理 5 确定广义逆算子 $\boldsymbol{A}^-$ 的特征性质为

$$\boldsymbol{AA}^-\boldsymbol{A}=\boldsymbol{A}.$$

推论 $\boldsymbol{GA}$ 是一个等幂算子.

因为
$$(\boldsymbol{GA})^2=\boldsymbol{GAGA}=\boldsymbol{GA}.$$

由此可知，$\boldsymbol{GA}$ 是一个投影算子. 实际上它是从酉空间 $\mathscr{C}^n$ 沿零空间 $\mathscr{N}(\boldsymbol{A})$ 到子空间 S 上的投影算子.

§3 相容方程的极小范数解和广义逆算子$\boldsymbol{A}_m^-$

3-1 相容方程的极小范数解和子空间 S 的选择

现在我们假定 $\boldsymbol{G}_S$ 是由子空间 S 所确定的算子 $\boldsymbol{A}$ 的一个广义逆算子 $\boldsymbol{A}^-$. 这时，相容方程(1)的任一解 x 都可唯一地表示成

$$\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v},$$

其中 $\boldsymbol{u}\in S$, $\boldsymbol{v}\in\mathscr{N}(\boldsymbol{A})$, 而 u 便是广义逆算子 $\boldsymbol{G}_S$ 所确定的解，即 $\boldsymbol{u}=\boldsymbol{G}_S\boldsymbol{b}$. 因为(1)的解的全体为线性流形

$$\boldsymbol{u}+\mathcal{N}(\boldsymbol{A}).$$

所以首先的问题就是求 $\boldsymbol{v}\in\mathcal{N}(\boldsymbol{A})$，使

$$\|\boldsymbol{x}\|=\|\boldsymbol{u}+\boldsymbol{v}\|=\text{极小}. \tag{10}$$

为此目的，我们把 $\boldsymbol{x}$ 表示成两个直交向量的和

$$\boldsymbol{x}=(\boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u})+(\boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}), \tag{11}$$

因为 $\boldsymbol{P}_{\mathcal{N}(A)}$ 是从空间 $\mathscr{C}^n$ 到 $\mathcal{N}(\boldsymbol{A})$ 上的直交投影算子，所以

$$\boldsymbol{P}_{\mathcal{N}(A)}(\boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u})=\boldsymbol{0},$$

这就是说 $\qquad \boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}\in\mathcal{N}^{\perp}(A).$

又， $\qquad \boldsymbol{P}_{\mathcal{N}(A)}(\boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u})=\boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u},$

故知 $\qquad \boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}\in\mathcal{N}(\boldsymbol{A}).$

于是便证明了(11)右端两个向量的直交性，即

$$(\boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u})\perp(\boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}).$$

从而据(11)便有

$$\|\boldsymbol{x}\|^2=\|\boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}\|^2+\|\boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}\|^2. \tag{12}$$

因此，当 $\qquad \boldsymbol{v}+\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}=\boldsymbol{0}$

时，$\|\boldsymbol{x}\|$ 为极小．这时

$$\boldsymbol{x}=\boldsymbol{u}-\boldsymbol{P}_{\mathcal{N}(A)}\boldsymbol{u}=(\boldsymbol{I}-\boldsymbol{P}_{\mathcal{N}(A)})\boldsymbol{u}, \tag{13}$$

其中的 $\boldsymbol{I}$ 为从 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 的不变算子，而 $\boldsymbol{I}-\boldsymbol{P}_{\mathcal{N}(A)}$ 是 $\boldsymbol{P}_{\mathcal{N}(A)}$ 的补投影算子，即，它是从空间 $\mathscr{C}^n$ 到子空间 $\mathcal{N}^{\perp}(\boldsymbol{A})$上的直交投影算子．这就是说，相容方程(1)（对给定的右端 $\boldsymbol{b}$)的解 $\boldsymbol{x}$，当它属于 $\mathcal{N}^{\perp}(\boldsymbol{A})$时，它的范数为极小，反之亦然．

从上述讨论可知，欲使 $\boldsymbol{Gb}$ 对任意的 $\boldsymbol{b}$ 都具有极小范数，必须且只需选 $\mathcal{N}^{\perp}(\boldsymbol{A})$ 作为子空间 S 即可，这时空间 $\mathscr{C}^n$ 的直接和分解为

$$\mathscr{C}^n=\mathcal{N}^{\perp}(\boldsymbol{A})\oplus\mathcal{N}(\boldsymbol{A}) \tag{14}$$

这种空间分解所确定的算子 $\boldsymbol{A}$ 的广义逆算子 $\boldsymbol{G}$，我们把它记作 $\boldsymbol{A}_m^-$.

$$x = A_m^- b \tag{15}$$

便是相容方程(1)的极小范数解.

如果算子 A 为满秩，这时空间分解无选择余地，A^- 和 A_m^- 完全相同.

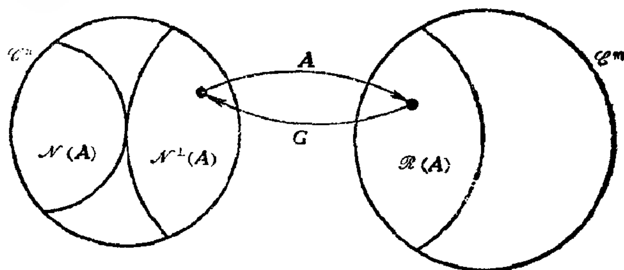


图 2-2　确定广义逆算子 A_m^- 的示意图

3-2　A_m^- 的特征性质

因为 A_m^- 同时也是一个 A^-，所以它也必满足关系

$$AA_m^-A = A. \tag{16}$$

此外，前面已经讲过，GA 是从空间 $\mathscr{C}^n$ 沿子空间 $\mathscr{N}(A)$ 到子空间 S 上的投影算子，即，它是一个等幂算子．对于 A_m^- 来说，由于确定空间分解的子空间为

$$S = \mathscr{N}^\perp(A),$$

所以 A_m^-A 是从 $\mathscr{C}^n$ 到 $\mathscr{N}^\perp(A)$ 上的直交投影算子.
这就是说

$$GA = P_{\mathscr{N}^\perp(A)} \tag{17}$$

因此，由第 1 章定理 13 知，GA 必为自伴算子，即

$$(GA)^* = GA. \tag{18}$$

反之，如果从 $\mathscr{R}(A)$ 到空间 $\mathscr{C}^n$ 的线性算子 G 满足条件

$$AGA = A, \quad (GA)^* = GA, \tag{19}$$

则 G 必是一个 A_m^-.

这里我们还要指出，算子 $\boldsymbol{G}$ 的定义域是 $\mathscr{R}(\boldsymbol{A})$，而并不是整个空间 $\mathscr{C}^m$. 对于广义逆算子 $\boldsymbol{A}_m^-$ 来说，由于空间分解(14)是唯一的，所以，作为从 $\mathscr{R}(\boldsymbol{A})$ 到 $\mathscr{C}^n$ 的线性算子来说，它是唯一的. 如果把 $\boldsymbol{G}$ 的定义域从 $\mathscr{R}(\boldsymbol{A})$ 扩张到整个空间 $\mathscr{C}^m$，这种唯一性就消失了.

§4 矛盾方程的最小二乘解和广义逆算子 $\boldsymbol{A}_l^-$

4-1 广义逆算子 $\boldsymbol{A}^-$ 的扩张

当算子方程(1)右端的向量 $\boldsymbol{b}$ 不属于 $\mathscr{R}(\boldsymbol{A})$ 时，方程(1)没有通常意义下的解. 在§1中已经讲过，这时可以考虑求 $\boldsymbol{x}$ 使

$$\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|=\text{极小} \tag{2}$$

的问题. 空间分解

$$\mathscr{C}^n=S\oplus\mathscr{N}(\boldsymbol{A}) \tag{5}$$

所确定的广义逆算子 $\boldsymbol{G}=\boldsymbol{A}^-$，其定义域为 $\mathscr{R}(\boldsymbol{A})$，而现在 $\boldsymbol{b}\notin\mathscr{R}(\boldsymbol{A})$，所以 $\boldsymbol{G}\boldsymbol{b}$ 无意义. 为了构造与问题(2)有关的广义逆算子 $\boldsymbol{G}$，首先需要把 $\boldsymbol{G}$ 的定义域从 $\mathscr{R}(\boldsymbol{A})$ 扩张到整个空间 $\mathscr{C}^m$. 为了这个目的，我们先对酉空间 $\mathscr{C}^m$ 作直接和分解

$$\mathscr{C}^m=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{U}, \tag{20}$$

其中 $\mathscr{U}$ 是 $\mathscr{C}^m$ 的一个子空间. 若 $\dim\mathscr{R}(\boldsymbol{A})=r$，则

$$\dim\mathscr{U}=m-r.$$

这时，$\mathscr{C}^m$ 中的任何向量 $\boldsymbol{b}$ 都可以唯一地表示成

$$\boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{b}_2,\quad \boldsymbol{b}_1\in\mathscr{R}(\boldsymbol{A}),\quad \boldsymbol{b}_2\in\mathscr{U}.$$

若设 $\boldsymbol{G}$ 为空间分解(5)所确定的一个广义逆算子 $\boldsymbol{A}^-$，则 $\boldsymbol{x}_1=\boldsymbol{G}\boldsymbol{b}_1$ 便是相容算子方程

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}_1 \tag{21}$$

在通常意义下的解，即

$$AGb_1 - b_1 = 0,$$

从而有

$$\|Ax_1 - b\| = \|Ax_1 - b_1 - b_2\| = \|b_2\|. \tag{22}$$

这里的 b_2 实际上就是 b 在空间分解(20)所确定的从 $\mathscr{C}^m$ 沿子空间 $\mathscr{R}(A)$ 到子空间 $\mathscr{U}$ 上的投影算子 $P_{\mathscr{U}}$ 映射下的象

$$b_2 = P_{\mathscr{U}} b \tag{23}$$

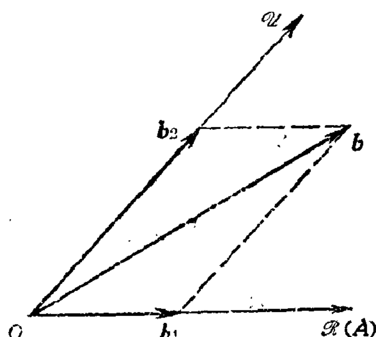


图 2-3

现在我们进一步来考虑为使 $\|b_2\|$ = 极小，应如何选择 $\mathscr{C}^m$ 的子空间 $\mathscr{U}$ 以构成空间的直接和分解. 当子空间 $\mathscr{U}$ 按这个原则确定以后，我们再把空间分解(5)所确定的广义逆算子 G 按下法加以扩张：

$$\text{对一切 } b = b_1 + b_2 \in \mathscr{C}^m, \text{ 令 } Gb = Gb_1, \tag{24}$$

这样扩张以后的广义逆算子 G，便是本节所需要的求问题(2)的解的广义逆算子.

现在问题的关键便是设法确定子空间 $\mathscr{U}$. 为了解决这个问题，我们可以从§3中讨论构造 A_m^- 相应的空间分解中得到启发.

首先，我们把向量 b_2 分解成两个直交向量之和：

$$\boldsymbol{b}_2 = (\boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}) + \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b} \tag{25}$$

因为 $\boldsymbol{b}_1 \in \mathscr{R}(\boldsymbol{A})$，所以 $\boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}_1 = \boldsymbol{0}$，从而可知

$$\boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b} = \boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}_2 = (\boldsymbol{I} - \boldsymbol{P}_{\mathscr{R}^\perp(A)})\boldsymbol{b}_2$$
$$= \boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}_2 \in \mathscr{R}(\boldsymbol{A}),$$

因此有

$$(\boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}) \perp \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}.$$

这就证明，(25)右端的两个向量互相直交．于是

$$\|\boldsymbol{b}_2\| = \|\boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}\|^2 + \|\boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}\|^2. \tag{26}$$

欲 $\|\boldsymbol{b}_2\|$ 为极小，从(26)显然可知，其充分必要条件为

$$\boldsymbol{b}_2 - \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b} = \boldsymbol{0},$$

即

$$\boldsymbol{b}_2 = \boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b}. \tag{27}$$

换句话说，$\boldsymbol{b}_2$ 应为向量 $\boldsymbol{b}$ 在 $\mathscr{R}^\perp(\boldsymbol{A})$ 上的直交投影.

根据这一段讨论可知，如果要想对任意给定的 $\boldsymbol{b} \in \mathscr{C}^m$，都使

$$\|\boldsymbol{AGb} - \boldsymbol{b}\| = \text{极小},$$

则为确定 $\boldsymbol{G}$ 应选子空间

$$\mathscr{U} = \mathscr{R}^\perp(\boldsymbol{A}), \tag{28}$$

并规定：当 $\boldsymbol{b} \in \mathscr{R}^\perp(\boldsymbol{A})$ 时

$$\boldsymbol{Gb} = \boldsymbol{0}.$$

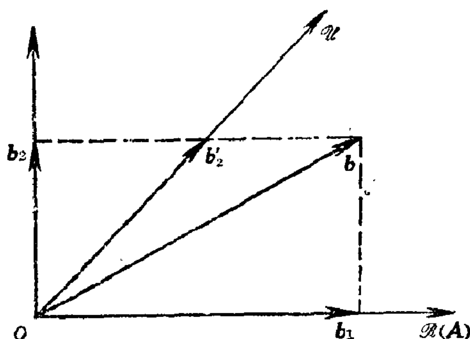


图 2-4

这样扩张后的算子 $\boldsymbol{G}$，它的定义域便是整个空间 $\mathscr{C}^m$，它就是我们所需要的广义逆算子.

上图说明前述讨论的几何意义. 对一般选择的子空间 $\mathscr{U}$, $\boldsymbol{b}$ 沿子空间 $\mathscr{R}(\boldsymbol{A})$ 到 $\mathscr{U}$ 上的投影为 $\boldsymbol{b}_2'$，而沿 $\mathscr{R}(\boldsymbol{A})$ 到 $\mathscr{R}^{\perp}(\boldsymbol{A})$ 上的投影为 $\boldsymbol{b}_2$，显然有 $\|\boldsymbol{b}_2\| \leqslant \|\boldsymbol{b}_2'\|$.

定义 6 按照上述方法所确定的线性算子 $\boldsymbol{A}$ 的广义逆算子 $\boldsymbol{G}$，记作 $\boldsymbol{A}_l^-$.

现在我们把确定算子 $\boldsymbol{G}=\boldsymbol{A}_l^-$ 的方法归纳于下：

第一步：给定空间 $\mathscr{C}^n$ 的直接和分解

$$\mathscr{C}^n = S \oplus \mathscr{N}(\boldsymbol{A}), \tag{5}$$

据此确定从 $\mathscr{R}(\boldsymbol{A})$ 到 S 的广义逆算子 $\boldsymbol{G}=\boldsymbol{A}^-$.

第二步：对空间 $\mathscr{C}^m$ 作直接和分解

$$\mathscr{C}^m = \mathscr{R}(\boldsymbol{A}) \oplus \mathscr{R}^{\perp}(\boldsymbol{A}); \tag{29}$$

把所得的广义逆算子 $\boldsymbol{G}$ 的定义域扩张到整个空间 $\mathscr{C}^m$ 上，规定

$$\boldsymbol{G}\boldsymbol{b} = \boldsymbol{G}(\boldsymbol{b}_1+\boldsymbol{b}_2) = \boldsymbol{G}\boldsymbol{b}_1; \quad \boldsymbol{b}_2 \in \mathscr{R}^{\perp}(\boldsymbol{A}), \quad \boldsymbol{b}_1 \in \mathscr{R}(\boldsymbol{A}). \tag{30}$$

这样就得到了广义逆算子 $\boldsymbol{A}_l^-$. 从前述过程可以看出，(5) 中的子空间 S 仍为一可变因素，它的选择不是唯一的，因而所得广义逆算子 $\boldsymbol{G}$ 也不是唯一的.

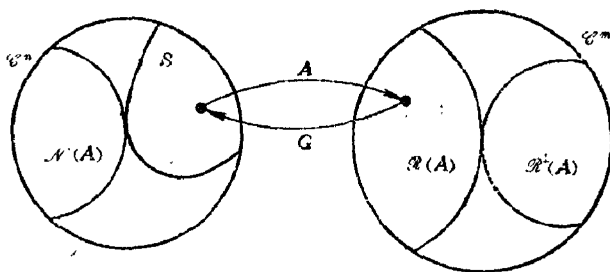


图 2-5 确定 $\boldsymbol{A}_l^-$ 的示意图

定理 6 广义逆算子 $G=A_l^-$ 是线性的.

证明 设 b_1, b_2 为 $\mathscr{C}^m$ 中任意两个向量. 令

$$\left.\begin{aligned}&b_i=b_1^{(i)}+b_2^{(i)},\\&b_1^{(i)}\in\mathscr{R}(A),\ b_2^{(i)}\in\mathscr{R}^\perp(A),\end{aligned}\right\}\ i=1,\ 2.$$

根据 G 的定义可知

$$Gb_i=Gb_1^{(i)}\in S,\ i=1,\ 2,$$

所以,对任意两个复数 α_1 和 α_2 恒有

$$G(\alpha_1 b_1+\alpha_2 b_2)=G(\alpha_1 b_1^{(1)}+\alpha_2 b_1^{(2)}),$$

因为 $\alpha_1 b_1^{(1)}+\alpha_2 b_1^{(2)}\in\mathscr{R}(A)$,已知在 $\mathscr{R}(A)$ 上 G 是线性的,所以

$$G(\alpha_1 b_1^{(1)}+\alpha_2 b_1^{(2)})=\alpha_1 Gb_1^{(1)}+\alpha_2 Gb_1^{(2)}=\alpha_1 Gb_1+\alpha_2 Gb_2$$

从而可知,定义在 $\mathscr{C}^m$ 上的算子 G 是线性的. 证毕.

4-2 法方程

根据 $G=A_l^-$ 的定义可知,若

$$b=b_1+b_2,\ b_1\in\mathscr{R}(A),\ b_2\in\mathscr{R}^\perp(A),$$

则

$$AGb=P_{\mathscr{R}(A)}b.$$

这就是说, $x_1=Gb$ 为方程(1)的最小二乘解的充分必要条件为

$$Ax_1=P_{\mathscr{R}(A)}b.$$

据第 1 章 5-3 公式(57),

$$\mathscr{R}^\perp(A)=\mathscr{N}(A^*),$$

所以

$$b_2=P_{\mathscr{N}(A^*)}b. \tag{27}$$

因此,若 x_1 为问题(2)的一个最小二乘解,则

$$A^*(b-Ax_1)=A^*b_2=0.$$

反之,若 x_1 满足方程

$$A^*(b-Ax_1)=0,$$

把向量 b 分解为

$$\boldsymbol{b}=\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}+\boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b},$$

则有

$$\boldsymbol{A}^*(\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_1+\boldsymbol{P}_{\mathscr{R}^\perp(A)}\boldsymbol{b})=\boldsymbol{A}^*(\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_1)=\boldsymbol{0},$$

但
$$\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}_1\in\mathscr{R}(\boldsymbol{A}),$$

所以必有
$$\boldsymbol{A}\boldsymbol{x}_1=\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b}.$$

这就证明 $\boldsymbol{x}_1$ 是问题(2)的一个最小二乘解.

根据以上讨论，我们有

定理 7 欲 $\boldsymbol{x}$ 为问题(2)的一个最小二乘解，其充分必要条件为

$$\boldsymbol{A}^*(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})=\boldsymbol{0}. \tag{31}$$

定义 7 我们称(31)为线性最小二乘问题(2)的法方程.

4-3 广义逆算子 $\boldsymbol{A}_l^-$ 的特征性质

前面所定义的广义逆算子 $\boldsymbol{G}=\boldsymbol{A}_l^-$，它也是一个 $\boldsymbol{A}^-$，所以必满足关系

$$\boldsymbol{AGA}=\boldsymbol{A}.$$

此外，由于对任何 $\boldsymbol{b}\in\mathscr{C}^m$，恒有

$$\boldsymbol{AGb}=\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b},$$

所以，乘积算子 $\boldsymbol{AG}$ 实际上就是从 $\mathscr{C}^m$ 到子空间 $\mathscr{R}(\boldsymbol{A})$ 上的直交投影算子. 因此，$\boldsymbol{AG}$ 不但具有等幂性

$$(\boldsymbol{AG})^2=\boldsymbol{AGAG}=\boldsymbol{AG}, \tag{32}$$

而且还必须具有自伴性，即

$$(\boldsymbol{AG})^*=\boldsymbol{AG}. \tag{33}$$

反之，对于任何一个从 $\mathscr{C}^m$ 到 $\mathscr{C}^n$ 内的线性算子 $\boldsymbol{G}$，如果它满足关系(32)～(33)，则 $\boldsymbol{G}$ 必为一个 $\boldsymbol{A}_l^-$.

定理 8 从 $\mathscr{C}^m$ 到 $\mathscr{C}^n$ 内的线性算子 $\boldsymbol{G}$ 是一个 $\boldsymbol{A}_l^-$ 的充分必要条件为(32)和(33)成立.

关系(32)和(33)就是广义逆算子 $\boldsymbol{A}_l^-$ 的特征性质.

§5 矛盾方程的极小最小二乘解和广义逆算子 A^+

5-1 空间分解的选择和广义逆算子 A^+

设线性算子 G 为空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 的直接和分解

$$\mathscr{C}^n = S \oplus \mathscr{N}(A)$$

$$\mathscr{C}^m = \mathscr{R}(A) \oplus \mathscr{R}^\perp(A)$$

所确定的广义逆算子 A_l^-. 如果

$$\operatorname{rank}(A) = n,$$

这时 $\dim \mathscr{N}(A) = 0$, 从而

$$S = \mathscr{C}^n.$$

从而子空间 S 无其它选择余地. 否则的话, S 的选择有无限多种, 而不同的选择, 所确定的广义逆算子 A_l^- 也各不相同, 因而, 对固定的向量 b, 随着子空间 S 的不同选择, 最小二乘解 Gb 也不同. 但是, 根据前面的讨论, 不难看出, $Ax-b$ 的最小范数总为

$$\|P_{\mathscr{R}^\perp(A)} b\|.$$

另一方面, 由 G 所得到的问题(2)的这个最小二乘解 $x = Gb$, 实际上是相容方程

$$Ax = P_{\mathscr{R}^\perp(A)} b \tag{34}$$

的解. 反之亦然. 所以, 求问题(2)的最小二乘解中具有极小范数的解的问题, 实际上就化为求相容方程(34)的极小范数解的问题. 根据§3中的讨论可知, 只有选取子空间 $\mathscr{N}^\perp(A)$ 作为 S, 才能使确定的广义逆算子 G, 对任意固定的 $b \in \mathscr{C}^m$, Gb 都具有极小范数.

根据以上的讨论, 我们便得出结论: 由空间分解

$$\mathscr{C}^n = \mathscr{N}(A) \oplus \mathscr{N}^\perp(A), \tag{35}$$

$$\mathscr{C}^m=\mathscr{R}(A)\oplus\mathscr{R}^{\perp}(A),$$

所确定的线性算子 A 的广义逆算子 G，就具有所需要的性质，即对任何 $b\in\mathscr{C}^m$，Gb 都是方程

$$Ax=b$$

的极小最小二乘解．算子 A 的这种广义逆算子，记之为 A^+．

这里应着重指出，因为确定广义逆算子 A^+ 的空间分解(35)是唯一的，所以 A^+ 也是唯一的．

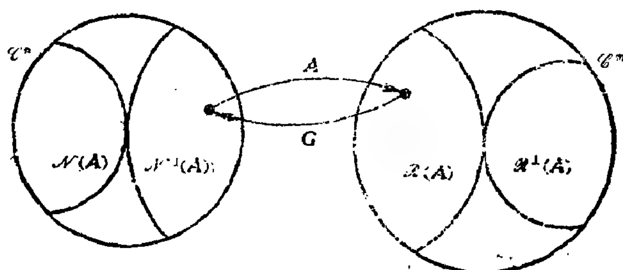


图 2-6 确定广义逆算子 A^+ 的示意图

5-2 广义逆算子 A^+ 的特征性质

仍用 G 来表示算子 A 的广义逆算子 A^+．因为，对于算子方程 $Ax=b$，Gb 是它的最小二乘解，即，它是一个 A_l^-，所以

$$AG=P_{\mathscr{R}(A)},$$

同时，Gb 也是它的极小范数解，也是一个 A_m^-，故有

$$GA=P_{\mathscr{R}(G)}.$$

于是有

定理 9 若 G 是确定算子方程极小最小二乘解的广义逆算子 A^+，则下列两个条件必同时成立：

$$AG=P_{\mathscr{R}(A)},\quad GA=P_{\mathscr{R}(G)}.\tag{36}$$

附注 实际上 $P_{\mathscr{R}(G)}$ 是把 $\mathscr{C}^n$ 映射到

$$\mathcal{N}^{\perp}(\boldsymbol{A})=\mathcal{R}(\boldsymbol{A}^*)$$

上的直交投影算子，所以可把它表示成

$$\boldsymbol{P}_{\mathcal{R}(G)}=\boldsymbol{P}_{\mathcal{R}(A^*)}. \tag{37}$$

此外还需要注意，现在，$\boldsymbol{G}$ 是从 $\mathscr{C}^m$ 到 $\mathscr{C}^n$ 内的一个线性算子，其定义域为整个空间 $\mathscr{C}^m$.

(36)就是通常所说的 Moore 条件，它还可以表示成下面的等价条件.

定理 10 Moore 条件(36)等价于下列条件:

$$\begin{gathered}\boldsymbol{AGA}=\boldsymbol{A},\ \boldsymbol{GAG}=\boldsymbol{G},\\(\boldsymbol{AG})^*=\boldsymbol{AG},\ (\boldsymbol{GA})^*=\boldsymbol{GA},\end{gathered} \tag{38}$$

这四个条件便是通常所说的 Penrose 条件.

证明 1) 若 Moore 条件(36)成立，则 Penrose 条件(38)也成立.

因为 $\boldsymbol{AG}=\boldsymbol{P}_{\mathcal{R}(A)}$，所以

$$\boldsymbol{AGA}=(\boldsymbol{AG})\boldsymbol{A}=\boldsymbol{P}_{\mathcal{R}(A)}\boldsymbol{A}=\boldsymbol{A}.$$

而从 $\boldsymbol{GA}=\boldsymbol{P}_{\mathcal{R}(G)}$，可立刻推出

$$\boldsymbol{GAG}=(\boldsymbol{GA})\boldsymbol{G}=\boldsymbol{P}_{\mathcal{R}(G)}\boldsymbol{G}=\boldsymbol{G}.$$

又由于 $\boldsymbol{AG}$ 和 $\boldsymbol{GA}$ 都是直交投影算子，所以都是自伴算子，即

$$(\boldsymbol{AG})^*=\boldsymbol{AG},\quad (\boldsymbol{GA})^*=\boldsymbol{GA}.$$

2) 反之，若 Penrose 条件(38)成立，则 Moore 条件也必成立.

$\boldsymbol{AG}$, $\boldsymbol{GA}$ 都是等幂算子，因为

$$\begin{gathered}(\boldsymbol{AG})^2=\boldsymbol{AGAG}=\boldsymbol{AG},\\(\boldsymbol{GA})^2=\boldsymbol{GAGA}=\boldsymbol{GA}.\end{gathered}$$

所以，$\boldsymbol{AG}$ 和 $\boldsymbol{GA}$ 都是投影算子. 根据 Penrose 的后两个条件，它们都是自伴算子，从而也都是直交投影算子. 从而可

知，Moore 条件(36)同时成立．证毕．

最后我们指出，若 $\mathrm{rank}(\boldsymbol{A})=n$，因为 $\boldsymbol{GA}=\boldsymbol{P}_{\mathscr{R}(G)}=P_{\mathscr{R}(A^*)}=\boldsymbol{P}_{\mathscr{C}^n}$，所以，对一切 $\boldsymbol{x}\in\mathscr{C}^n$，恒有

$$\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x},$$

故必有

$$\boldsymbol{A}^+\boldsymbol{A}=\boldsymbol{I}_n. \tag{39}$$

而若 $\mathrm{rank}(\boldsymbol{A})=m$，则

$$\boldsymbol{A}\boldsymbol{A}^+=\boldsymbol{I}_m, \tag{40}$$

其中的 $\boldsymbol{I}_n$ 和 $\boldsymbol{I}_m$ 分别表示空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 上的不变算子．

5-3 $\boldsymbol{A}^+$ 的若干性质

1）有关秩的一些性质

1° $\mathrm{rank}(\boldsymbol{A}^+)=\mathrm{rank}(\boldsymbol{A})$;

2° $\mathrm{rank}(\boldsymbol{A}^+\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^+)=\mathrm{rank}(\boldsymbol{A})$.

证明 因为 $\mathscr{R}(\boldsymbol{A}^+)=\mathscr{N}^\perp(\boldsymbol{A})=\mathscr{R}(\boldsymbol{A}^*)$，故有

1° $\dim\mathscr{R}(\boldsymbol{A}^+)=\dim\mathscr{R}(\boldsymbol{A}^*)=\dim\mathscr{R}(\boldsymbol{A})=\mathrm{rank}(\boldsymbol{A})$.

2° $\mathscr{R}(\boldsymbol{A}^+\boldsymbol{A})=\mathscr{R}(\boldsymbol{A}^+)$，所以

$$\dim\mathscr{R}(\boldsymbol{A}^+\boldsymbol{A})=\dim\mathscr{R}(\boldsymbol{A}^+)=\mathrm{rank}(\boldsymbol{A}).$$

同理有 $\dim\mathscr{R}(\boldsymbol{A}\boldsymbol{A}^+)=\mathrm{rank}(\boldsymbol{A})$.

2）根据 Penrose 条件可直接验证下列关系式成立：

3° $(\alpha\boldsymbol{A})^+=\alpha^{-1}\boldsymbol{A}^+$, $\alpha\neq 0$;

4° $(\boldsymbol{A}^*)^+=(\boldsymbol{A}^+)^*$;

5° $(\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}^*)^+=\boldsymbol{V}\boldsymbol{A}^+\boldsymbol{U}^*$, $\boldsymbol{U}$, $\boldsymbol{V}$ 分别为 $\mathscr{C}^m$ 和 $\mathscr{C}^n$ 上的酉算子．

3）乘积关系

6° $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{A}^*\boldsymbol{A}\boldsymbol{A}^+=\boldsymbol{A}^*$;

7° $\boldsymbol{A}\boldsymbol{A}^*\boldsymbol{A}^{*+}=\boldsymbol{A}^{*+}\boldsymbol{A}^*\boldsymbol{A}=\boldsymbol{A}$.

证明 $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^*=\boldsymbol{P}_{\mathscr{R}(A^*)}\boldsymbol{A}^*=\boldsymbol{A}^*$.

又，　　$(A^*AA^+)^* = A^{*+}A^*A = P_{\mathscr{R}(A)}A = A,$

所以　　$$A^*AA^+ = A^*.$$

故 6° 得证，把 6° 两端作运算(*)便可得到 7°.

8°　$(A^*A)^+ = A^+A^{*+}$;

9°　$(AA^*)^+ = A^{*+}A^+$.

根据 Penrose 条件并利用关系 6°～7°，可直接验证.

4) 满秩算子的广义逆

10°　若 $\operatorname{rank}(A) = n$，则

$$A^+ = (A^*A)^{-1}A^*.$$

11°　若 $\operatorname{rank}(A) = m$，则

$$A^+ = A^*(AA^*)^{-1}.$$

这里的 $(A^*A)^{-1}$ 和 $(AA^*)^{-1}$ 分别为从 $\mathscr{C}^n \to \mathscr{C}^n$, $\mathscr{C}^m \to \mathscr{C}^m$ 上的满秩线性算子 A^*A 和 AA^* 的逆算子，即

$$(A^*A)^{-1}(A^*A) = (A^*A)(A^*A)^{-1} = I_n,$$

$$(AA^*)^{-1}(AA^*) = (AA^*)(AA^*)^{-1} = I_m.$$

证明　据 6°,

$$A^*AA^+ = A^*,$$

因为 A^*A 有逆算子 $(A^*A)^{-1}$，用它左乘上式两端便得到 10°. 同理，据 7° 有

$$AA^*A^{*+} = A,$$

而 AA^* 有逆算子 $(AA^*)^{-1}$，用它左乘上式两端并作运算*便得到 11°. 证毕.

特例　设 A 为从 $\mathscr{C}^1$ 到 $\mathscr{C}^n$ 的线性算子，且

$$\operatorname{rank}(A) = 1,$$

则　　$$A^+ = (A^*A)^{-1}A^*.$$

若 A 为从 $\mathscr{C}^n$ 到 $\mathscr{C}^1$ 的线性算子，则

$$A^+ = A^*(AA^*)^{-1}.$$

在第一种情形，因为 A^*A 是一个从 $\mathscr{C}^1$ 到 $\mathscr{C}^1$ 的线性算子，它是一个复数，而由于 $\operatorname{rank}(A^*A)=\operatorname{rank}(A)=1$，所以 $A^*A\neq 0$. 在第二种情形，AA^* 也是一个异于零的复数. 在建立空间坐标系以后，我们将要看到第一种情形的算子可以用一个 n 维列向量来表示，而第二种情形的算子 A，则可以用一个 n 维的行向量来表示. 在两种情形，A^*A 以及 AA^* 都是相应向量的范数平方.

5-4 乘积 AB 的广义逆算子

在线性代数中，我们已经知道，若 A, B 是两个 $n\times n$ 阶非奇异方阵，则有公式

$$(AB)^{-1}=B^{-1}A^{-1}.$$

对于广义逆算子来说，公式

$$(AB)^{+}=B^{+}A^{+} \tag{41}$$

是否仍然成立？

如果 $B=A^*$，或 $A=B^*$，根据前面的性质 8°～9°，可知公式(41)成立. 若 $\operatorname{rank}(A)=\operatorname{rank}(B)=r$，而 A 和 B 分别为从 $\mathscr{C}^r$ 到 $\mathscr{C}^m$，和从 $\mathscr{C}^n$ 到 $\mathscr{C}^r$ 上的线性算子，这时，由于

$$A^{+}=(A^*A)^{-1}A^*,\ B^{+}=B^*(BB^*)^{-1},$$

不难验证乘积公式(41)也成立. 但是，在一般情形，公式(41)不成立. 例如，若取 $A=[1,\ 0]$, $B=[1,\ 1]^T$，则易知

$$(AB)^{+}=1,\ B^{+}A^{+}=\frac{1}{2},$$

所以二者不等.

这就说明乘积公式(41)有时成立，有时又不成立. 现在我们来考虑公式(41)成立的充分必要条件是什么？现在已经知道，这种充要条件可以表示成多种形式，在本节中我们只举出其中的两种.

今设 $\boldsymbol{A}$, $\boldsymbol{B}$ 分别为从 $\mathscr{C}^r$ 到 $\mathscr{C}^m$ 和从 $\mathscr{C}^n$ 到 $\mathscr{C}^r$ 的线性算子，则 $\boldsymbol{AB}$ 为从 $\mathscr{C}^n$ 到 $\mathscr{C}^m$ 的线性算子.

定理 11 乘积公式

$$(\boldsymbol{AB})^+=\boldsymbol{B}^+\boldsymbol{A}^+ \tag{41}$$

成立的一种充分必要条件为:

(i) $\boldsymbol{A}^+\boldsymbol{ABB}^*\boldsymbol{A}^*=\boldsymbol{BB}^*\boldsymbol{A}^*$

(ii) $\boldsymbol{BB}^+\boldsymbol{A}^*\boldsymbol{AB}=\boldsymbol{A}^*\boldsymbol{AB}$

同时成立.

证明 设条件(i)(ii)都成立. 用 $\boldsymbol{B}^+$ 和$(\boldsymbol{AB})^{*+}$分别左、右乘 $\boldsymbol{A}^+\boldsymbol{ABB}^*\boldsymbol{A}^*$ 得

$$\boldsymbol{B}^+\boldsymbol{A}^+\boldsymbol{ABB}^*\boldsymbol{A}^*(\boldsymbol{AB})^{*+}=\boldsymbol{B}^+\boldsymbol{A}^+(\boldsymbol{AB})(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}$$

据性质 7° 有

$$(\boldsymbol{AB})(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}=\boldsymbol{AB}.$$

所以 $$\boldsymbol{B}^+\boldsymbol{A}^+(\boldsymbol{AB})(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}=\boldsymbol{B}^+\boldsymbol{A}^+\boldsymbol{AB}.$$

另一方面

$$\begin{aligned}\boldsymbol{B}^+(\boldsymbol{BB}^*\boldsymbol{A}^*)(\boldsymbol{AB})^{*+}&=(\boldsymbol{B}^+\boldsymbol{BB}^*)\boldsymbol{A}^*(\boldsymbol{AB})^{*+}\\&=\boldsymbol{B}^*\boldsymbol{A}^*(\boldsymbol{AB})^{*+}=(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}.\end{aligned}$$

于是，据$(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}$的自伴性得

$$\boldsymbol{B}^+\boldsymbol{A}^+\boldsymbol{AB}=(\boldsymbol{AB})^*(\boldsymbol{AB})^{*+}=(\boldsymbol{AB})^+(\boldsymbol{AB}). \tag{a}$$

对(ii)两端再作运算*,

$$\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{AB}^{+*}\boldsymbol{B}^*=\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{A}.$$

用 $\boldsymbol{A}^+$, $(\boldsymbol{AB})^{*+}$ 分别右、左乘两端得

$$\begin{aligned}(\boldsymbol{AB})^{*+}\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{AB}^{+*}\boldsymbol{B}^*\boldsymbol{A}^+&=(\boldsymbol{AB})^{*+}(\boldsymbol{AB})^*\boldsymbol{AB}^{*+}\boldsymbol{B}^*\boldsymbol{A}^+\\&=(\boldsymbol{AB})^{*+}(\boldsymbol{AB})^*\boldsymbol{A}(\boldsymbol{BB}^+)^*\boldsymbol{A}^+\\&=(\boldsymbol{AB})^{*+}(\boldsymbol{AB})^*(\boldsymbol{AB})\boldsymbol{B}^+\boldsymbol{A}^+=\boldsymbol{ABB}^+\boldsymbol{A}^+.\end{aligned}$$

又 $$\begin{aligned}(\boldsymbol{AB})^{*+}\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{AA}^+&=(\boldsymbol{AB})^{*+}\boldsymbol{B}^*\boldsymbol{A}^*\\&=(\boldsymbol{AB})^{*+}(\boldsymbol{AB})^*=(\boldsymbol{AB})(\boldsymbol{AB})^+.\end{aligned}$$

所以

$$ABB^+A^+=(AB)(AB)^+. \tag{b}$$

(a)和(b)分别可以表示成投影算子

(iii) $$B^+A^+AB=P_{\mathscr{R}((AB)^*)},$$

(iv) $$ABB^+A^+=P_{\mathscr{R}(AB)}.$$

这与 B^+A^+ 为 $(AB)^+$ 的 Moore 条件相似. 据此，我们可以证明乘积公式成立(参看附录 2).

反过来，假定乘积公式成立. 因为

$$(AB)^+(AB)(AB)^*=(AB)^*,$$

$$B^*A^*=B^+A^+ABB^*A^*.$$

左乘以 ABB^*B, 据 $B^*BB^+=B^*$ 得

$$ABB^*BB^*A^*=AB(B^*BB^+)A^+ABB^*A^*$$

$$=ABB^*A^+ABB^*A^*.$$

所以 $$ABB^*(I-A^+A)BB^*A^*=0.$$

因为 $I-A^+A$ 为等幂算子，从而上式可以表示成

$$ABB^*(I-A^+A)^2BB^*A^*$$

$$=[ABB^*(I-A^+A)][ABB^*(I-A^+A)]^*=0.$$

从而可知 $$(I-A^+A)BB^*A^*=0.$$

这就证明条件(i)成立. 仿此可以证明条件(ii)也成立. 证毕.

根据这个定理可以证明

定理 12 乘积公式(41)成立的充分必要条件为:

1° $BB^*\mathscr{R}(A^*)\subset\mathscr{R}(A^*)$,

2° $A^*A\mathscr{R}(B)\subset\mathscr{R}(B)$.

证明 设条件 1° 成立，则对任何向量 x,

$$BB^*A^*x\in\mathscr{R}(A^*).$$

而 $A^+A=P_{\mathscr{R}(A^*)}$，所以

$$A^+A(BB^*A^*)x=P_{\mathscr{R}(A^*)}(BB^*A^*x)=(BB^*A^*)x.$$

于是便证明条件(i)成立.

同理,若 2° 成立,可证条件(ii)也成立.

这就证明条件 1°、2° 是充分的.

反之,若条件(i)、(ii)成立,今证 1° 和 2° 也必成立.

条件(i)可以写成,对任意 x

$$P_{\mathscr{R}(A^*)}BB^*A^*x=BB^*A^*x,$$

所以
$$BB^*\mathscr{R}(A^*)\subset\mathscr{R}(A^*).$$

此外,条件(ii)可以表示成,对任何 x

$$P_{\mathscr{R}(B)}A^*ABx=A^*ABx,$$

可知
$$A^*A\mathscr{R}(B)\subset\mathscr{R}(B).$$

这就是说条件 1°、2° 都成立.因此 1° 和 2° 也是必要的.证毕.

§6 广义逆算子的矩阵表示

6-1 广义逆算子 A^- 的矩阵表示

本节总假定 A 是从酉空间 $\mathscr{C}^n$ 到 $\mathscr{C}^m$ 内的一个线性算子.广义逆算子 A^- 是由空间 $\mathscr{C}^n$ 的直接和分解

$$\mathscr{C}^n=\mathscr{N}(A)\oplus S \tag{42}$$

所确定的,它的定义域是 $\mathscr{C}^m$ 中的一个子空间 $\mathscr{R}(A)$,其值域为子空间 S.子空间 S 和 $\mathscr{R}(A)$的维数相同,设

$$\dim\mathscr{R}(A)=\dim S=r. \tag{43}$$

这也就是说,算子 A 把子空间 $S\subset\mathscr{C}^n$ 映射成子空间 $\mathscr{R}(A)\subset\mathscr{C}^m$,而 A^- 则把子空间 $\mathscr{R}(A)$映射成 S:

$$S\underset{A^-}{\overset{A}{\rightleftarrows}}\mathscr{R}(A) \tag{44}$$

或者说,A 在 S 上的象空间是 $\mathscr{R}(A)$,而 A^- 的象空间为 S,

$$AS=\mathscr{R}(A),\quad A^{-}\mathscr{R}(A)=S.$$

若 $x\in S$, $y=Ax$, 则 $A^{-}y=x$. 当子空间 S 给定以后, A^{-} 不仅是存在的, 而且也是唯一的. 在子空间 S 和 $\mathscr{R}(A)$ 中分别建立坐标基底后, 我们可以按照以前所讲过的方法把 A^{-} 用一个 $r\times r$ 阶矩阵表示出来. 设

$$e_1,\ e_2,\ \cdots,\ e_r \text{ 和 } g_1,\ g_2,\ \cdots,\ g_r \tag{45}$$

分别为子空间 S 和子空间 $\mathscr{R}(A)$ 的一组基底. 设

$$\begin{aligned}
A^{-}g_1&=\alpha_{11}e_1+\alpha_{21}e_2+\cdots+\alpha_{r1}e_r,\\
A^{-}g_2&=\alpha_{12}e_1+\alpha_{22}e_2+\cdots+\alpha_{r2}e_r,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
A^{-}g_r&=\alpha_{1r}e_1+\alpha_{2r}e_2+\cdots+\alpha_{rr}e_r.
\end{aligned}\tag{46}$$

并设 $y\in\mathscr{R}(A)$, $x=A^{-}y$, 它们在前述基底下的表示为:

$$x_e=\begin{bmatrix}\xi_1\\ \vdots\\ \xi_r\end{bmatrix},\quad y_g=\begin{bmatrix}\eta_1\\ \vdots\\ \eta_r\end{bmatrix},$$

即

$$\begin{aligned}
x&=\xi_1e_1+\xi_2e_2+\cdots+\xi_re_r,\\
y&=\eta_1g_1+\eta_2g_2+\cdots+\eta_rg_r.
\end{aligned}$$

于是我们便有关系

$$\begin{aligned}
x=A^{-}y&=A^{-}(\eta_1g_1+\eta_2g_2+\cdots+\eta_rg_r)=\sum_{i=1}^{r}\eta_iA^{-}g_i\\
&=\sum_{j=1}^{r}\left(\sum_{i=1}^{r}\eta_i\alpha_{ji}\right)e_j=\xi_1e_1+\xi_2e_2+\cdots+\xi_re_r.
\end{aligned}$$

令两端 e_i 的系数相等便得到关系

$$\xi_j=\sum_{i=1}^{r}\eta_i\alpha_{ji},\quad j=1,\ 2,\ \cdots,\ r.$$

从而可知

$$x_e=\begin{bmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_r\end{bmatrix}=\begin{bmatrix}\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1r}\\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2r}\\ \cdots & \cdots & \cdots & \cdots\\ \alpha_{r1} & \alpha_{r2} & \cdots & \alpha_{rr}\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_r\end{bmatrix}. \tag{47}$$

把矩阵 $[\alpha_{ij}]$ 记作 G_r, 它就是在给定基底下广义逆算子 $\boldsymbol{A}^-$ 的矩阵表示, 或记之为

$$A^-_{e,g}=G_r.$$

在上述坐标系中, 线性算子 $\boldsymbol{A}$ (定义域看成是 S) 也同样可以表示成一个 $r\times r$ 阶矩阵, 记之为 $A_{e,g}$, 则显然 $A_{e,g}$ 非奇异, 且

$$G_r=A^-_{e,g}.$$

线性算子 $\boldsymbol{A}$ 的定义域实际上是 $\mathscr{C}^n$, 其值域在 $\mathscr{C}^m$ 内. 在空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中都建立坐标系以后, 可以用一个 $m\times n$ 阶矩阵把它表示出来. 而广义逆算子 $\boldsymbol{A}^-$ 的定义域仅为 $\mathscr{C}^m$ 中的一个子空间 $\mathscr{R}(\boldsymbol{A})$, 其值域为 $\mathscr{C}^n$ 中的子空间 S, 要把它表示成矩阵, 需要在有关的子空间中选出基底来, 这并不是很容易的事. 另一方面, 在实际应用上, 我们直接考虑的对象往往是线性算子的矩阵表示. 表示算子 $\boldsymbol{A}$ 的矩阵为 $m\times n$ 阶, 如果把 $\boldsymbol{A}^-$ 的定义域扩张到整个 $\mathscr{C}^m$, 那么, $\boldsymbol{A}^-$ 便可以用一个 $n\times m$ 阶矩阵表示出来, 现在我们在原则上给出扩张定义域的方法.

6-2 广义逆算子 $\boldsymbol{A}^-$ 的扩张和广义逆矩阵 A^-

为了把广义逆算子 $\boldsymbol{A}^-$ 的定义域从 $\mathscr{R}(\boldsymbol{A})$ 扩张到整个空间 $\mathscr{C}^m$, 并设法用一个 $n\times m$ 阶矩阵把它表示出来, 首先对空间 $\mathscr{C}^m$ 作直接和分解

$$\mathscr{C}^m=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{U}, \tag{48}$$

其中 $\dim\mathscr{U}=m-r$. 在空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中分别取

$$\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n \text{ 和 } \boldsymbol{g}_1, \cdots, \boldsymbol{g}_m$$

作为坐标基底, 为便于讨论, 并假定:

$$\begin{aligned}&\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_r \text{ 为 } S \text{ 的基底},\\&\boldsymbol{e}_{r+1}, \cdots, \boldsymbol{e}_n \text{ 为 } \mathscr{N}(\boldsymbol{A}) \text{ 的基底},\end{aligned} \tag{49}$$

$$\boldsymbol{g}_1, \boldsymbol{g}_2, \cdots, \boldsymbol{g}_r \text{ 为 } \mathscr{R}(\boldsymbol{A}) \text{的基底，}$$
$$\boldsymbol{g}_{r+1}, \boldsymbol{g}_{r+2}, \cdots, \boldsymbol{g}_m \text{ 为 } \mathscr{U} \text{ 的基底.}$$

今规定

$$\boldsymbol{A}^- \boldsymbol{y} = \boldsymbol{A}^-(\boldsymbol{y}_1 + \boldsymbol{y}_2) = \boldsymbol{A}^- \boldsymbol{y}_1, \quad (\boldsymbol{y}_1 \in \mathscr{R}(\boldsymbol{A}), \boldsymbol{y}_2 \in \mathscr{U}.) \tag{50}$$

这就是说，对任何向量 $\boldsymbol{u} \in \mathscr{U}$，恒有 $\boldsymbol{A}^- \boldsymbol{u} = \boldsymbol{0}$. 这样就把 $\boldsymbol{A}^-$ 的定义域扩张到了整个空间 $\mathscr{C}^m$ 上，并且扩张后的算子仍为线性算子.

在上述基底下，向量 $\boldsymbol{x} \in S$，$\boldsymbol{y} \in \mathscr{R}(\boldsymbol{A})$ 的坐标向量具有以下特殊的形式：

$$\boldsymbol{x}_e = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{y}_g = \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{51}$$

根据用矩阵表示线性算子的一般方法，$\boldsymbol{A}^-$ 便可以用一个 $n \times m$ 阶矩阵表示出来，今记之为 $A^-_{n\times m}$. 它必须具有下列性质：

$$A^-_{n\times m} \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_m \end{bmatrix} = \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_r \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

而且，$\xi_1, \xi_2, \cdots, \xi_r$ 与 $\eta_{r+1}, \eta_{r+2}, \cdots, \eta_m$ 无关. 因此，在前述坐标系的特殊选择下，$A^-_{n\times m}$ 的形式为

$$\begin{bmatrix} G_r & 0 \\ 0 & 0 \end{bmatrix}.$$

如果在 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中分别选取 $\{\boldsymbol{e}'_i\}$ 和 $\{\boldsymbol{g}'_j\}$ 作为坐标基底，在这两个坐标系中向量 x 和 y 的坐标向量记为 $\boldsymbol{x}_{e'}$ 和 $\boldsymbol{y}_{g'}$. 设新旧坐标向量之间的关系为

$$\boldsymbol{x}_e = P\boldsymbol{x}_{e'}, \quad \boldsymbol{y}_g = Q\boldsymbol{y}_{g'},$$

则在新的坐标系中，$A^-_{n\times m}$ 的形式为

$$P\begin{bmatrix} G_r & 0 \\ 0 & 0 \end{bmatrix}Q,$$

其中 P 和 Q 分别为 $n\times n$ 阶和 $m\times m$ 阶非奇异矩阵，乘积是一个 $n\times m$ 阶矩阵.

现在我们假定，在空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中建立了坐标系以后，线性算子 $\boldsymbol{A}$ 和它的广义逆算子 $\boldsymbol{A}^-$ 的矩阵表示分别为 A 和 A^-，我们便称 A^- 为矩阵 A 的一种**广义逆矩阵**. 由于线性算子乘积的矩阵表示，为其各算子矩阵表示之积，所以刻划广义逆矩阵 A^- 的特征性质便可以表示为

$$AA^-A = A. \tag{52}$$

若 $\boldsymbol{x}$ 和 $\boldsymbol{b}$ 都是在所述坐标系中的坐标向量，$A^-\boldsymbol{b}$ 便是相容线性方程组

$$A\boldsymbol{x} = \boldsymbol{b}$$

的一个解.

6-3　广义逆算子 $\boldsymbol{A}_m^-$ 的矩阵表示和广义逆矩阵 A_m^-

对于广义逆算子 $\boldsymbol{A}_m^-$，可以通过与前述完全一样的方法用矩阵表示出来. 这时，由于确定算子 $\boldsymbol{A}_m^-$ 的空间分解为

$$\mathscr{C}^n = \mathscr{N}(\boldsymbol{A}) \oplus \mathscr{N}^\perp(\boldsymbol{A}), \tag{53}$$

前述子空间 S 只能取为 $\mathscr{N}^\perp(\boldsymbol{A})$，无其它选择. 如果分别在子空间 $\mathscr{N}^\perp(\boldsymbol{A})$ 和 $\mathscr{R}(\boldsymbol{A})$ 中建立坐标系，$\boldsymbol{A}_m^-$ 就可以唯一地表示成一个 $r\times r$ 阶的矩阵 G_r. 这个矩阵还是非奇异的. 如果仿前把算子 $\boldsymbol{A}_m^-$ 的定义域扩张到整个空间 $\mathscr{C}^m$ 上，即对 $\mathscr{C}^m$

作空间分解

$$\mathscr{C}^m=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{U}, \tag{54}$$

子空间 $\mathscr{U}$ 是 $\mathscr{R}(\boldsymbol{A})$ 的一个补空间，其维数为 $m-r$，它有一定的任意性．这样，只对 $\mathscr{U}$ 中的向量 $\boldsymbol{u}$，规定

$$\boldsymbol{A}_m^-(\boldsymbol{y}_1+\boldsymbol{y}_2)=\boldsymbol{A}_m^-\boldsymbol{y}_1,\quad \boldsymbol{y}_1\in\mathscr{R}(\boldsymbol{A}),\boldsymbol{y}_2\in\mathscr{U},$$

就完成了算子 $\boldsymbol{A}_m^-$ 的扩张． 如果在空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中建立坐标系，算子 $\boldsymbol{A}_m^-$ 就可以唯一地表示成 $n\times m$ 阶矩阵，记之为 A_m^-．因为算子 $\boldsymbol{A}_m^-$ 具有以下的特征性质：

$$\left.\begin{aligned}\boldsymbol{A}\boldsymbol{A}_m^-\boldsymbol{A}&=\boldsymbol{A},\\(\boldsymbol{A}_m^-\boldsymbol{A})^*&=\boldsymbol{A}_m^-\boldsymbol{A},\end{aligned}\right\} \tag{55}$$

故 $\boldsymbol{A}_m^-$ 的矩阵表示 A_m^- 也具有相应的特征性质：

$$\left.\begin{aligned}AA_m^-A&=A,\\(A_m^-A)^*&=A_m^-A,\end{aligned}\right\} \tag{56}$$

这里的记号 A，A_m^- 都是相应算子的矩阵表示，它们分别为 $m\times n$ 阶和 $n\times m$ 阶． 记号$(^*)$在(55)中表示求伴随算子，而在(56)中表示共轭转置运算．

算子 $\boldsymbol{A}_m^-\boldsymbol{A}$ 是 $\mathscr{C}^n$ 到 $\mathscr{N}(\boldsymbol{A})$上的直交投影算子，它的矩阵表示 A_m^-A 便是相应的投影矩阵．对于以 A 为系数矩阵的线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 来说，若 $\boldsymbol{b}\in\mathscr{R}(\boldsymbol{A})$，$\boldsymbol{x}=A_m^-\boldsymbol{b}$ 便是它的极小范数解．

6-4 广义逆算子 $\boldsymbol{A}_l^-$ 的矩阵表示和广义逆矩阵 A_l^-

表示矛盾算子方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的最小二乘解的广义逆算子 $\boldsymbol{A}_l^-$ 决定于下列空间分解

$$\left.\begin{aligned}\mathscr{C}^n&=\mathscr{N}(\boldsymbol{A})\oplus S,\\\mathscr{C}^m&=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{R}^\perp(\boldsymbol{A}),\end{aligned}\right\} \tag{57}$$

和 $\boldsymbol{A}^-$ 的情形一样，子空间 S 也有一定的任意性，而空间 $\mathscr{C}^m$ 的分解是完全确定了的，所述算子方程的最小二乘解，实际上

就是相容算子方程

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{P}_{\mathscr{R}(\boldsymbol{A})}\boldsymbol{b} \tag{58}$$

的解，即我们有关系

$$\boldsymbol{A}_l^-=\boldsymbol{A}^-\boldsymbol{P}_{\mathscr{R}(\boldsymbol{A})}. \tag{59}$$

由此可以看出，广义逆算子 $\boldsymbol{A}_l^-$，它的定义域是整个空间 $\mathscr{C}^m$，实际上仅为广义逆算子(定义域仅为 $\mathscr{R}(\boldsymbol{A})$) $\boldsymbol{A}^-$ 的一种按特殊方式的扩张：

$$\left.\begin{array}{c}\boldsymbol{A}_l^-\boldsymbol{b}=\boldsymbol{A}^-(\boldsymbol{b}_1+\boldsymbol{b}_2)=\boldsymbol{A}_l^-\boldsymbol{b}_1,\\ \text{其中 } \boldsymbol{b}=\boldsymbol{b}_1+\boldsymbol{b}_2,\ \boldsymbol{b}_1\in\mathscr{R}(\boldsymbol{A}),\ \boldsymbol{b}_2\in\mathscr{R}^\perp(\boldsymbol{A}).\end{array}\right\} \tag{60}$$

和前面一样，在空间 $\mathscr{C}^n$ 和 $\mathscr{C}^m$ 中建立坐标系以后，算子 $\boldsymbol{A}_l^-$ 可以用一个 $n\times m$ 阶矩阵表示出来，记之为 A_l^-. 对于以表示算子 $\boldsymbol{A}$ 的 $m\times n$ 阶矩阵 A 为系数矩阵的线性方程组

$$Ax=b$$

来说，A_l^-b 就是它的一个最小二乘解. 我们称 A_l^- 为矩阵 A 的一个确定最小二乘解的广义逆矩阵. 和它所表示的广义逆算子 $\boldsymbol{A}_l^-$ 一样，广义逆矩阵 A_l^- 具有下列特征性质：

$$\begin{aligned}&AA_l^-A=A,\\&(AA_l^-)^*=AA_l^-,\end{aligned} \tag{61}$$

其中，AA_l^- 为直交投影算子 $\boldsymbol{A}\boldsymbol{A}_l^-=\boldsymbol{P}_{\mathscr{R}(\boldsymbol{A})}$ 的矩阵表示，所以我们称之为一个直交投影矩阵.

6-5 广义逆算子 $\boldsymbol{A}^+$ 的矩阵表示和广义逆矩阵 A^+

线性算子 $\boldsymbol{A}$ 的广义逆算子 $\boldsymbol{A}^+$，为下列空间分解所唯一确定：

$$\left.\begin{array}{l}\mathscr{C}^n=\mathscr{N}(\boldsymbol{A})\oplus\mathscr{N}^\perp(\boldsymbol{A}),\\ \mathscr{C}^m=\mathscr{R}(\boldsymbol{A})\oplus\mathscr{R}^\perp(\boldsymbol{A}),\end{array}\right\} \tag{62}$$

对于线性算子方程 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 来说，$\boldsymbol{A}^+\boldsymbol{b}$ 是它的唯一的极小最小二乘解. 和 $\boldsymbol{A}_l^-$ 的情形相仿，$\boldsymbol{A}^+$ 也可以看成是广义逆

算子 $\boldsymbol{A}_{\bar{m}}^{-}$ 的一种特殊的扩张. 实际上, $\boldsymbol{A}^{+}\boldsymbol{b}$ 就是相容算子方程

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{P}_{\mathscr{R}(A)}\boldsymbol{b} \tag{63}$$

的极小范数解, 从而可知广义逆算子 $\boldsymbol{A}^{+}$ 还可以表示成

$$\boldsymbol{A}^{+}=\boldsymbol{A}_{m}^{-}\boldsymbol{P}_{\mathscr{R}(A)}, \tag{64}$$

应当指出, 这里的广义逆算子 $\boldsymbol{A}_{\bar{m}}^{-}$ 的定义域仍为 $\mathscr{R}(\boldsymbol{A})$. 在空间 $\mathscr{C}^{n}$, $\mathscr{C}^{m}$ 中分别建立坐标系后, 算子 $\boldsymbol{A}$, $\boldsymbol{A}^{+}$ 都可以分别表示成 $m\times n$ 阶矩阵 A 和 $n\times m$ 阶矩阵 A^{+}. 对于以 A 为系数矩阵的线性方程组

$$A\boldsymbol{x}=\boldsymbol{b}$$

来说, $A^{+}\boldsymbol{b}$ 便是它的极小最小二乘解, 所以称 A^{+} 为矩阵 A 的与相应问题求解的广义逆矩阵. 根据线性算子和它的矩阵表示之间的关系, A^{+} 具有下列特征性质:

$$\left.\begin{array}{ll} AA^{+}A=A, & A^{+}AA^{+}=A^{+}, \\ (AA^{+})^{*}=AA^{+}, & (A^{+}A)^{*}=A^{+}A, \end{array}\right\} \tag{65}$$

或

$$AA^{+}=P_{\mathscr{R}(A)},\ A^{+}A=P_{\mathscr{R}(A^{*})}, \tag{66}$$

这里的 $P_{\mathscr{R}(A)}$ 和 $P_{\mathscr{R}(A^{*})}$ 分别为直交投影算子 $\boldsymbol{P}_{\mathscr{R}(A)}$ 和 $\boldsymbol{P}_{\mathscr{R}(A^{*})}$ 的矩阵表示, 即它们是相应的直交投影矩阵.

我们称 A^{+} 为矩阵 A 的 **Moore-Penrose 广义逆.**

§7 一些特殊矩阵的 Moore-Penrose 广义逆

本节我们将列出一些特殊矩阵的 Moore-Penrose 广义逆, 由于证明都很容易, 故从略.

我们假定在今后所考虑的酉空间中都已选定了标准直交基底, 建立了相应的直角坐标系, 所提到的矩阵都是相应的线性算子在这些坐标系中的矩阵表示.

1° 若 A 是一个 $n\times n$ 阶非奇异方阵，则

$$A^+=A^{-1}.$$

2° $0^+_{m\times n}=0_{n\times m}$，即，$m\times n$ 阶零矩阵的广义逆是 $n\times m$ 阶的零矩阵.

3° 设 $D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$ 为 $n\times n$ 阶对角阵，则

$$D^+=\mathrm{diag}(d_1^+, d_2^+, \cdots, d_n^+),$$

其中记号 d_i^+ 表示

$$d_i^+=\begin{cases} 0, & 若\ d_i=0, \\ d_i^{-1}, & 若\ d_i\neq 0, \end{cases} \quad i=1, 2, \cdots, n.$$

4° $$\begin{bmatrix} A \\ 0 \end{bmatrix}^+=\begin{bmatrix} A^+ & 0 \end{bmatrix}.$$

5° 设 $A=USV^*$，A 为 $m\times n$ 阶，U 和 V 分别为 $m\times m$ 阶和 $n\times n$ 阶酉矩阵，而 S 为 $m\times n$ 阶矩阵，则

$$A^+=VS^+U^*.$$

6° 若 A 为 $m\times n$ 阶，$\mathrm{rank}(A)=n$，则

$$A^+=(A^*A)^{-1}A^*.$$

若 $\mathrm{rank}(A)=m$，则

$$A^+=A^*(AA^*)^{-1}.$$

7° 特例：$\boldsymbol{a}$ 为 n 维列向量，且 $\boldsymbol{a}\neq\boldsymbol{0}$，则

$$\boldsymbol{a}^+=(\boldsymbol{a}^*\boldsymbol{a})^{-1}\boldsymbol{a}^*,$$

而

$$\boldsymbol{a}^{*+}=\boldsymbol{a}^{+*}=\boldsymbol{a}(\boldsymbol{a}^*\boldsymbol{a})^{-1}.$$

8° 满秩分解．设 $A=BC$，A 为 $m\times n$ 阶，B 为 $m\times r$ 阶，C 为 $r\times n$ 阶，且

$$\mathrm{rank}(A)=\mathrm{rank}(B)=\mathrm{rank}(C)=r,$$

则

$$A^+=C^+B^+.$$

作为本章的结束语，我们补充说明于下.

1）在前面的讨论中，在建立各种广义逆算子及其特征性

质时，有关空间为酉空间，重要的是其中定义了内积，并不涉及具体的坐标系。换句话说，向量是抽象的向量，而不是坐标向量，算子是抽象的算子，它们都与坐标系的选择无关。仅仅在建立矩阵表示的时候，才需要建立坐标系。

2）在本章中所考虑的线性算子为定义在酉空间上的线性算子，它们的矩阵表示为复矩阵。如果把酉空间换成欧氏空间，那么相应的矩阵就是实矩阵了。这时，$\boldsymbol{A}$ 的伴随算子 $\boldsymbol{A}^*$ 可表示成 $\boldsymbol{A}^T$，而相应的伴随矩阵

$$A^* = A^T.$$

这就是说，把共轭转置（*）换成转置（^{T}）就得到在欧氏空间中的相应表达式。

第 3 章

广义逆矩阵的摄动理论和连续性问题

§1 矩阵的奇异值分解和奇异值摄动定理

1-1 Courant-Fischer 定理

关于广义逆矩阵以及和它相联系的线性最小二乘问题，有一大类来自于对统计的或实验观测数据处理的需要. 问题中的数据本身带有一定的误差. 在利用电子计算机进行计算的过程中，由于计算机的字长为有限，不可避免地还要产生舍入误差. 这两种误差对计算结果都会产生影响. 我们所得到的计算结果，实际上并不是原始问题的解，而是原始问题的某一摄动问题的解. 摄动量的大小决定于原始数据的精度以及所使用计算方法的数值稳定性. 在本章的讨论中，我们暂且不管这种摄动的来源，而只从理论上来研究一定的摄动量对解所产生的影响，并指出为避免或减小摄动的严重危害，对摄动的大小应控制到何种程度. 这就是我们研究摄动理论的任务和目的. 在本章中，我们将就实矩阵的情形进行讨论，但是，所得到的结论以及所使用的方法，对于复矩阵的情形也同样适用. 在当前这一小节中，我们首先向读者介绍反映实对称矩阵特征值极性的 Courant-Fischer 定理，为今后的讨论提供所需要的基础. 在以后的讨论中，我们总假定在所涉及的欧氏空间中，已经建立了一个直角坐标系，也就是说，已经选定了一个标准直交系作为空间的基底.

为了讨论实对称矩阵特征值的极性，我们先引进关于实

对称矩阵的谱分解公式.

设 A 是一个 $n\times n$ 阶的实对称矩阵，则它的 n 个特征值都是实数，并存在一个完备的标准直交特征向量系. 今假定这 n 个实特征值按大小顺序排列为:

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n \tag{1}$$

和它们相应的标准直交特征向量系为:

$$\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \cdots,\ \boldsymbol{u}_n.$$

即，

$$A\boldsymbol{u}_i=\lambda_i\boldsymbol{u}_i,\ \boldsymbol{u}_i^T\boldsymbol{u}_j=\begin{cases}1,\ i=j,\\0,\ i\neq j,\end{cases}\quad i,\ j=1,\ 2,\ \cdots,\ n. \tag{2}$$

定理 1 任何 $n\times n$ 阶的实对称矩阵 A，总可以用它的一个标准直交特征向量系 $\{\boldsymbol{u}_i\}$ 分解成单秩矩阵之和:

$$A=\lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T+\cdots+\lambda_n\boldsymbol{u}_n\boldsymbol{u}_n^T. \tag{3}$$

分解式(3)就叫做实对称矩阵 A 的谱分解公式.

证明 用 A 的特征向量 $\boldsymbol{u}_j$ 为列构造矩阵

$$U=[\boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \cdots,\ \boldsymbol{u}_n],$$

显然 U 是一个直交阵，并且

$$\begin{aligned}U^TAU&=U^T[A\boldsymbol{u}_1,\ \cdots A\boldsymbol{u}_n]\\&=U^T[\lambda_1\boldsymbol{u}_1,\ \cdots,\ \lambda_n\boldsymbol{u}_n]\\&=\Lambda=\mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n).\end{aligned} \tag{4}$$

从而便得到谱分解公式

$$A=U\Lambda U^T=\lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T+\cdots+\lambda_n\boldsymbol{u}_n\boldsymbol{u}_n^T.$$ 证毕.

现在我们来考虑下列极值问题:

问题 1 任意给定 r 个 n 维向量 $\boldsymbol{p}_1,\ \boldsymbol{p}_2,\ \cdots\boldsymbol{p}_r$，求

$$\max_{\substack{\|x\|=1\\ x^Tp_i=0}}\boldsymbol{x}^TA\boldsymbol{x}. \tag{5}$$

然后，对一切包含 r 个 n 维向量的集合 $\{\boldsymbol{p}_i\}$，求

$$\min_{\{p_i\}} \max_{\substack{\|x\|=1 \\ x^T p_i=0}} x^T A x. \tag{6}$$

问题 2 在问题 1 的同样条件下，先求

$$\min_{\substack{\|x\|=1 \\ x^T p_i=0}} x^T A x$$

再变动 $\{p_i\}$ 求

$$\max_{\{p_i\}} \min_{\substack{\|x\|=1 \\ x^T p_i=0}} x^T A x. \tag{7}$$

这就是本节所要考虑的二次型的 min-max 和 max-min 问题.

今就问题 1 讨论于下.

因为 $U^T A U = \Lambda$，所以，若令

$$x = Uy, \text{ 则 } y = U^T x,$$

于是便有

$$x^T A x = y^T \Lambda y = \lambda_1 \eta_1^2 + \cdots + \lambda_n \eta_n^2,$$

其中 $\eta_1, \eta_2, \cdots, \eta_n$ 为向量 y 的分量. 这时令

$$p_i = U q_i, \quad \text{即} \quad q_i = U^T p_i, \ i = 1, 2, \cdots, r.$$

于是条件 $\|x\|=1$, $x^T p_i = 0$ 就转化为

$$y^T y = x^T x = 1,$$

$$x^T p_i = y^T U^T p_i = y^T q_i = 0, \quad i = 1, 2, \cdots, r.$$

向量 y 的 n 个分量 $\eta_1, \cdots, \eta_n$ 满足 $r+1$ 个方程，其中有 r 个为齐次方程，

$$y^T q_i = 0, \ i = 1, 2, \cdots, r.$$

再补充 $n-r-1$ 个条件

$$\eta_{r+2} = \eta_{r+3} = \cdots = \eta_n = 0,$$

便得到含 n 个未知量 $\eta_1, \cdots, \eta_n$ 的 $n-1$ 个齐次方程，它必有非零解，设其为

$$\eta_1, \eta_2, \cdots, \eta_{r+1}, 0, \cdots, 0,$$

为使其满足另一个条件 $\|\boldsymbol{y}\|=1$，可以把它标准化

$$\sum_{i=1}^{r+1}\eta_i^2=1.$$

这时显然有 $\quad \boldsymbol{x}^T A\boldsymbol{x}=\boldsymbol{y}^T\Lambda\boldsymbol{y}=\sum_{i=1}^{r+1}\lambda_i\eta_i^2\geqslant\lambda_{r+1},$

从而可知，对前述任意给定的 r 个向量 $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_r$，必有

$$\max_{\substack{\|\boldsymbol{x}\|=1\\ \boldsymbol{x}^T\boldsymbol{p}_i=0}} \boldsymbol{x}^T A\boldsymbol{x}\geqslant\lambda_{r+1}. \tag{8}$$

由于 $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_r$ 是任意的，这就表示

$$\min_{(\boldsymbol{p}_i)}\max_{\substack{\|\boldsymbol{x}\|=1\\ \boldsymbol{x}^T\boldsymbol{p}_i=0}} \boldsymbol{x}^T A\boldsymbol{x}\geqslant\lambda_{r+1}.$$

另一方面，如果取 $\boldsymbol{p}_i=U\boldsymbol{e}_i$，这时 $\boldsymbol{q}_i=\boldsymbol{e}_i$, $i=1, 2, \cdots, r$，而条件 $\boldsymbol{y}^T\boldsymbol{y}=1$, $\boldsymbol{y}^T\boldsymbol{q}_i=0$ 就化为

$$\boldsymbol{y}^T\boldsymbol{e}_i=0,\ i=1, 2, \cdots, r,$$

即 $$\eta_i=0,\ i=1, 2, \cdots, r,$$

以及 $$\sum_{i=r+1}^{n}\eta_i^2=1.$$

因此，$\quad \boldsymbol{x}^T A\boldsymbol{x}=\boldsymbol{y}^T\Lambda\boldsymbol{y}=\sum_{i=r+1}^{n}\lambda_i\eta_i^2\leqslant\lambda_{r+1}.$

由此可见必有

$$\max_{\substack{\|\boldsymbol{x}\|=1\\ \boldsymbol{x}^T\boldsymbol{p}_i=0}} \boldsymbol{x}^T A\boldsymbol{x}\leqslant\lambda_{r+1}. \tag{9}$$

比较(8)和(9)可知，对如此给定的 $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_r$ 必有

$$\max_{\substack{\|\boldsymbol{x}\|=1\\ \boldsymbol{x}^T\boldsymbol{p}_i=0}} \boldsymbol{x}^T A\boldsymbol{x}=\lambda_{r+1}.$$

实际上这个极大值当

$$\eta_{r+1}=1,\ \eta_i=0,\ i\neq r+1,\ i=1, 2, \cdots, n$$

时达到. 所以，对一切包含 r 个向量 $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_r$ 的集合 $\{\boldsymbol{p}_i\}$，我们有关系

$$\min_{\{p_i\}} \max_{\substack{\|x\|=1\\ x^T p_i=0}} x^T A x=\lambda_{r+1}. \tag{10}$$

这就是问题 1 的解. 关于问题 2, 可以用同样的方法来讨论, 其结果为

$$\max_{\{p_i\}} \min_{\substack{\|x\|=1\\ x^T p_i=0}} x^T A x=\lambda_{n-r}. \tag{11}$$

于是便得到

定理 2 (Courant-Fischer)

$$\min_{\{p_i\}} \max_{\substack{\|x\|=1\\ x^T p_i=0}} x^T A x=\lambda_{r+1} \tag{10}$$

$$\max_{\{p_i\}} \min_{\substack{\|x\|=1\\ x^T p_i=0}} x^T A x=\lambda_{n-r}, \tag{11}$$

其中$\{p_i\}$包含有 r 个向量 $p_1, \cdots, p_r$.

1-2 矩阵的奇异值分解定理

实对称方阵可以用直交相似变换化为对角形, 其中对角阵的主对角元为矩阵的特征值. 对于一般的长方阵(也包括不对称的方阵)有非常相似的结论. 今设 A 是一个 $m\times n$ 阶实矩阵, 我们首先把它扩张成一个 $(m+n)\times(m+n)$ 阶的实对称方阵

$$C=\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \tag{12}$$

由于 C 为实对称矩阵, 所以存在有完备的标准直交特征向量系. 设 $w\in\mathscr{R}^{m+n}$ 是矩阵 C 的与特征值 λ 相应的特征向量

$$Cw=\lambda w. \tag{13}$$

$m+n$ 维向量 w 可以用一个 m 维向量 u 和一个 n 维向量 v 表示成

$$w=\begin{bmatrix} u \\ v \end{bmatrix}.$$

于是(13)就可以表示成

$$\begin{aligned} A\boldsymbol{v} &= \lambda \boldsymbol{u}, \\ A^T\boldsymbol{u} &= \lambda \boldsymbol{v}. \end{aligned} \tag{14}$$

用 A^T 和 A 分别乘这两个方程便得到

$$\begin{aligned} A^TA\boldsymbol{v} &= \lambda A^T\boldsymbol{u} = \lambda^2\boldsymbol{v}, \\ AA^T\boldsymbol{u} &= \lambda A\boldsymbol{v} = \lambda^2\boldsymbol{u}. \end{aligned} \tag{15}$$

这就说明 $\boldsymbol{v}$ 和 $\boldsymbol{u}$ 分别是正半定矩阵 A^TA 和 AA^T 的特征向量，它们相应的特征值为 λ^2. 显然 $\lambda(\geqslant 0)$ 和 $-\lambda$ 同时都是矩阵 C 的特征值，和它们相应的特征向量分别为

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \quad 和 \quad \begin{bmatrix} -\boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix}.$$

如果 $\operatorname{rank}(A)=k<n$，则 $\operatorname{rank}(C)=2k$，所以矩阵 C 恰有 $2k$ 个异于零的特征值，每一个特征值都成对出现把它们记为

$$\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_k \quad 和 \quad -\lambda_1,\ -\lambda_2,\ \cdots,\ -\lambda_k.$$

不失一般性，可假定 $\lambda_i>0,\ i=1,\ 2,\ \cdots,\ k.$

定义 1 我们称 $\lambda_1, \cdots, \lambda_k$ 为矩阵 A 的奇异值.

据(15)可知，A 和 A^T 有共同的奇异值.

如果 A 为列满秩，即 $\operatorname{rank}(A)=n$，则 A 有 n 个奇异值。有时为了方便，在 $\operatorname{rank}(A)=k$ 时，我们也称 A 有 $n-k$ 个零奇异值.

现在我们进一步假定正半定矩阵 A^TA 和 AA^T 的特征值 $\lambda_1^2,\ \lambda_2^2,\ \cdots, \lambda_k^2$ 相对应的特征向量分别为

$$\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots,\ \boldsymbol{v}_k \quad 和 \quad \boldsymbol{u}_1,\ \boldsymbol{u}_2,\ \cdots,\ \boldsymbol{u}_k.$$

从而有

$$\begin{aligned} A^TA[\boldsymbol{v}_1,\ \cdots,\ \boldsymbol{v}_k] &= [\lambda_1^2\boldsymbol{v}_1,\ \cdots,\ \lambda_k^2\boldsymbol{v}_k], \\ AA^T[\boldsymbol{u}_1,\ \cdots,\ \boldsymbol{u}_k] &= [\lambda_1^2\boldsymbol{u}_1,\ \cdots,\ \lambda_k^2\boldsymbol{u}_k]. \end{aligned} \tag{16}$$

因为

$$A[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k]=[A\boldsymbol{v}_1, \cdots, A\boldsymbol{v}_k]=[\lambda_1\boldsymbol{u}_1, \cdots, \lambda_k\boldsymbol{u}_k],$$

于是便得到了分解公式

$$AV=U\Lambda, \tag{17}$$

其中

$$\begin{gathered} U=[\boldsymbol{u}_1, \cdots, \boldsymbol{u}_k],\ V=[\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k], \\ \Lambda=\operatorname{diag}(\lambda_1, \cdots, \lambda_k). \end{gathered} \tag{18}$$

U, V 和 Λ分别为 $m\times k$ 阶, $n\times k$ 阶和 $k\times k$ 阶. 不失一般性, 我们可假定$\{\boldsymbol{u}_i\}$和$\{\boldsymbol{v}_i\}$都是标准直交系, 于是公式(17)就可以表示成

$$U^TAV=\Lambda,$$

或

$$A=U\Lambda V^T. \tag{19}$$

定义 2 我们称(19)为矩阵 A 的**奇异值分解公式.**

在这个分解式中, U 和 V 的列向量分别为矩阵 AA^T 和 A^TA 的特征向量, 而对角阵 Λ 中的主对角元素就是矩阵A 的奇异值. 适当交换矩阵 A 各列的次序, 可以使 Λ 中的奇异值 $\lambda_1, \lambda_2, \cdots, \lambda_k$ 为不增序列.

1-3 矩阵 A 和广义逆矩阵 A^+ 的谱范数

设 $m\times n$ 阶矩阵 A 的秩为 k, 则 A 有 k 个正奇异值

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k>0.$$

现在我们用奇异值来把矩阵及其广义逆矩阵的范数表示出来.

1° 矩阵 A 的范数 根据矩阵范数的定义,

$$\|A\|=\max_{\|\boldsymbol{x}\|=1}\|A\boldsymbol{x}\|, \ \boldsymbol{x}\in\mathscr{R}^n.$$

根据向量欧氏范数的表达式

$$\|A\boldsymbol{x}\|^2=\boldsymbol{x}^TA^TA\boldsymbol{x}.$$

矩阵 A^TA 为正半定, 它的 k 个正特征值依序为

$$\lambda_1^2 \geqslant \lambda_2^2 \geqslant \cdots \geqslant \lambda_k^2.$$

所以

$$\max_{\|x\|=1} \|Ax\|^2 = \max_{\|x\|=1} (x^T A^T A x) = \max_{\|x\|=1} \left(x^T \sum_{i=1}^{k} \lambda_i^2 v_i v_i^T \right) x$$

$$= \max_{\|x\|=1} \sum_{i=1}^{k} \lambda_i^2 (x^T v_i)^2.$$

不难证明 $$\sum_{i=1}^{k} (x^T v_i)^2 \leqslant 1,$$

从而可知 $$\max_{\|x\|=1} \|Ax\|^2 \leqslant \lambda_1^2.$$

然而，若令 $x = v_1$，则显然有

$$x^T A^T A x = \lambda_1^2,$$

从而可知 $$\max_{\|x\|=1} \|Ax\|^2 = \lambda_1^2,$$

所以我们便证明了

$$\|A\| = \lambda_1. \tag{20}$$

这就是说，矩阵 A 的范数 $\|A\|$ 等于正半定矩阵 $A^T A$ 的最大特征值的正平方根，或 $\|A\|$ 等于矩阵 A 的最大奇异值. 显然，如果 $A=0$，则 $\|A\|=0$. 这时零矩阵没有正奇异值，也可以说它的最大奇异值为零.

2° 广义逆矩阵 A^+ 的范数 根据矩阵 A 的奇异值分解公式，若 $\mathrm{rank}(A)=k$，则 $A=U\Lambda V^T$，U 为 $m\times k$ 阶，Λ 为 $k\times k$ 阶，V 为 $n\times k$ 阶，并且对角阵 Λ 非奇异. 所以

$$A^+ = V\Lambda^{-1}U^T. \tag{21}$$

我们仍假定 Λ 主对角元处的奇异值仍按不增顺序排列，并且

$$\Lambda^{-1} = \begin{bmatrix} \lambda_1^{-1} & & \\ & \ddots & \\ & & \lambda_k^{-1} \end{bmatrix}. \tag{22}$$

根据范数的定义

$$\|A^+\|^2=\max_{\|x\|=1}\|A^+x\|^2$$
$$=\max_{\|x\|=1}\{(V\Lambda^{-1}U^Tx)^T(V\Lambda^{-1}U^Tx)\}$$
$$=\max_{\|x\|=1}y^T\Lambda^{-2}y,$$

其中，$y=U^Tx$，由于 $x^Tx=1$，所以 $y^Ty=x^Tx=1$.

于是 $$\|A^+\|^2=\max_{\|y\|=1}y^T\Lambda^{-2}y=\lambda_k^{-2}.$$

这就得到

$$\|A^+\|=\lambda_k^{-1}, \tag{23}$$

即广义逆矩阵 A^+ 的范数等于矩阵 A 的最小(正)奇异值的倒数，也就是矩阵 A^+ 的最大奇异值.

因为矩阵 A 和 A^+ 的范数都可以用 A 的奇异值来表示，所以又称这种矩阵范数为矩阵的**谱范数**. 显然$\|0^+\|=\|0\|=0$.

1-4 奇异值摄动定理

今设 $m\times n$ 阶矩阵 A 有微小摄动 δA，记 $B=A+\delta A$，且

$$\mathrm{rank}(B)=\mathrm{rank}(A)=k.$$

记矩阵 A 的奇异值为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k,$$

摄动矩阵 B 由于与 A 同秩，也有 k 个正奇异值，

$$\sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_k.$$

设正半定矩阵 A^TA 的与特征值 $\lambda_1^2,\lambda_2^2,\cdots,\lambda_k^2$ 相应的特征向量为 $v_1, v_2, \cdots, v_k$. 取 Courant-Fischer 定理中的

$$p_i=v_i,\ i=1,\ 2,\ \cdots,\ r,\ r\leqslant k-1.$$

这时有

$$\sigma_{r+1}^2\leqslant\max_{\substack{\|x\|=1\\ x^Tp_i=0}}x^TB^TBx=\max_{\substack{\|x\|=1\\ x^Tp_i=0}}x^T(A+\delta A)^T(A+\delta A)x$$
$$\leqslant\max_{\substack{\|x\|=1\\ x^Tp_i=0}}\{(x^TA^TAx)^{\frac{1}{2}}+(x^T(\delta A)^T(\delta A)x)^{\frac{1}{2}}\}^2$$

$$\leqslant \{ \max_{\substack{\|x\|=1 \\ x^T p_i=0}} (x^T A^T A x)^{\frac{1}{2}} + \max_{\substack{\|x\|=1 \\ x^T p_i=0}} (x^T (\delta A)^T (\delta A) x)^{\frac{1}{2}} \}^2$$

$$\leqslant (\lambda_{r+1} + \|\delta A\|)^2, \quad r=1, 2, \cdots, k-1.$$

所以有

$$\sigma_{r+1} \leqslant \lambda_{r+1} + \|\delta A\|, \tag{24}$$

即有

$$\sigma_k \leqslant \lambda_k + \|\delta A\|.$$

于是我们就得到关于奇异值的摄动定理.

定理 3 设 A, δA, $B=A+\delta A$ 均为 $m\times n$ 阶矩阵, 且 $\mathrm{rank}(B)=\mathrm{rank}(A)$. 把 A 和 B 的 k 个奇异值按大小顺序分别排列为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k,$$

$$\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_k,$$

则以下的结论成立:

$$\sigma_i \leqslant \lambda_i + \|\delta A\|, \quad i=1, 2, \cdots, k. \tag{25}$$

§2 Banach 引理和非奇异方阵的摄动定理

2-1 引言

本章的主要目的在于研究广义逆矩阵 A^+ 的摄动理论,并据以解决广义逆矩阵的连续性问题. 为此, 我们先从非奇异方阵着手.

2-2 Banach引理

设 A 是一个 $n\times n$ 阶非奇异方阵, 即逆矩阵 A^{-1} 存在. 经微小摄动后, $A+\delta A$ 是否仍能保持它的非奇异性? 显然, 如果 λ 是 A 的一个特征值, 若取 $\delta A=-\lambda I$, 则 $A+\delta A$ 就变成奇异的了. 当然, 如果取 μ 异于 A 的所有特征值, 不论 μ 如何大, 矩阵 $A-\mu I$ 总是非奇异的. 但是, 我们感兴趣的是微小摄动 δA 对矩阵的影响, 也就是说, 是否当 $\|\delta A\|$ 足够小时,

恒能保持矩阵的非奇异性？若然，我们进一步就可以研究极限问题：

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^{-1}=?.$$

我们将证明一定存在正数 η，当 $\|\delta A\|\leqslant\eta$ 时，矩阵 $A+\delta A$ 恒为非奇异，并且，对任何非奇异方阵 A，极限关系

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^{-1}=A^{-1}$$

恒成立．这就是说，非奇异方阵 A 的逆矩阵 A^{-1} 是**连续**的．为此，我们先证明一个重要引理．

引理(Banach)　设 I 为 $n\times n$ 阶单位阵，F 是一个同阶方阵．若

$$\|F\|<1, \tag{26}$$

则矩阵 $I-F$ 非奇异，并且

$$\|(I-F)^{-1}\|\leqslant(1-\|F\|)^{-1}. \tag{27}$$

证明　设 $\boldsymbol{x}$ 为任一 n 维非零向量，总有

$$\|(I-F)\boldsymbol{x}\|=\|\boldsymbol{x}-F\boldsymbol{x}\|\geqslant\|\boldsymbol{x}\|-\|F\boldsymbol{x}\|\geqslant\|\boldsymbol{x}\|-\|F\|\|\boldsymbol{x}\|$$
$$=(1-\|F\|)\|\boldsymbol{x}\|>0.$$

因此，齐次方程 $(I-F)\boldsymbol{x}=0$ 无非零解，从而可知矩阵 $I-F$ 必为非奇异，即在条件(26)满足时，逆矩阵 $(I-F)^{-1}$ 存在．

现在我们再来证明不等式(27)．由于 $(I-F)^{-1}$ 存在，所以 $(I-F)(I-F)^{-1}=I$，这个恒等式可以表示成

$$(I-F)(I-F)^{-1}=(I-F)^{-1}-F(I-F)^{-1},$$

从而有

$$(I-F)^{-1}=I+F(I-F)^{-1}.$$

对两端取范数便得到不等式

$$\|(I-F)^{-1}\|\leqslant 1+\|F\|\|(I-F)^{-1}\|$$

即
$$\|(I-F)^{-1}\|(1-\|F\|)\leqslant 1,$$

从而便证明了不等式

$$\|(I-F)^{-1}\| \leqslant (1-\|F\|)^{-1}. \quad \text{证毕.}$$

这里指出，若以 $-F$ 代 F，则引理仍成立. 这就是说，若 $\|F\|<1$，则 $I+F$ 也为非奇异，并且

$$\|(I+F)^{-1}\| \leqslant (1-\|F\|)^{-1}.$$

推论 设 $\|F\|<1$，我们还有估计式

$$\|I-(I-F)^{-1}\| \leqslant \frac{\|F\|}{1-\|F\|}. \tag{28}$$

证明 因为这时存在逆矩阵 $(I-F)^{-1}$，并且

$$I-(I-F)^{-1} = -F(I-F)^{-1}.$$

两端取范数便得到

$$\|I-(I-F)^{-1}\| = \|F(I-F)^{-1}\| \leqslant \|F\| \|(I-F)^{-1}\|,$$

从而便证明了不等式(28). 证毕.

矩阵 $I^{-1}=I$，不等式 (28) 表示矩阵 I 摄动后的逆矩阵 $(I-F)^{-1}$ 逼近 I^{-1} 的程度，$\|F\|$ 愈小，两者愈为接近.

2-3 一般逆矩阵的摄动定理

上一小节中的结果是我们所要讨论的一般非奇异方阵的特例. 但是，有了这个基础，解决一般情形就没有困难了. 因为，若 $n\times n$ 阶方阵 A 为非奇异，则总可以把摄动矩阵 $A+\delta A$ 写成

$$A+\delta A = A(I+A^{-1}\delta A),$$

$A+\delta A$ 为非奇异的充分必要条件是 $(I+A^{-1}\delta A)$ 为非奇异. 并且当 $(I+A^{-1}\delta A)$ 为非奇异时，

$$(A+\delta A)^{-1} = (I+A^{-1}\delta A)^{-1}A^{-1}.$$

于是，一般情形的问题就转化为前面的特例，即现在的问题便是考虑保持矩阵 $I+A^{-1}\delta A$ 的非奇异性问题了. 令 $F=-A^{-1}\delta A$，这个问题就是 Banach 引理及其推论中所曾经讨论过的. 对于这种一般问题的结论，可以叙述如下.

定理 4 设 A 和 δA 为同阶方阵而 A 为非奇异. 若摄动

阵 δA 满足不等式

$$\|A^{-1}\delta A\|<1, \tag{29}$$

则下列结论成立:

1° $A+\delta A$ 为非奇异;

2° 估计式

$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|}\leqslant\frac{\|F\|}{1-\|F\|} \tag{30}$$

成立,其中 $F=A^{-1}\delta A$.

根据这个定理我们便可以证明

定理 5 非奇异方阵 A 的逆矩阵 A^{-1} 是连续的,即我们有极限关系

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^{-1}=A^{-1}.$$

证明 根据不等式(30),当 $\|\delta A\|\to 0$ 时,

$$\|F\|=\|A^{-1}\delta A\|\leqslant\|A^{-1}\|\|\delta A\|\to 0,$$

所以有

$$\lim_{\|\delta A\|\to 0}\|A^{-1}-(A+\delta A)^{-1}\|=0,$$

即

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^{-1}=A^{-1}.$$

此为欲证.

§3 满秩矩阵广义逆的连续性问题

3-1 广义逆矩阵不连续的例

前节中我们讨论了非奇异方阵逆矩阵的连续性问题. 结论是 A^{-1} 具有所述的连续性. 现在我们要问,保证这种连续性的根本原因何在?从表面上我们可以看到以下现象. 非奇异方阵为满秩,当摄动量 δA 足够小时,不会改变这个事实. 如果 A 不满秩,不难证明总可以选取 δA,一方面可使之任意小,而同时使 $A+\delta A$ 变为满秩,即 $(A+\delta A)^{-1}$ 存在. 另一方

面，微小摄动也可以保持 $A+\delta A$ 为奇异，从而就不能考虑极限问题 $\lim\limits_{\|\delta A\|\to 0}(A+\delta A)^{-1}$. 实际上，不论如何推广逆矩阵的概念，有关连续性问题的答案都是否定的. 今以广义逆矩阵 A^+ 为例作为说明，因为 A^+ 的存在不受 $\operatorname{rank}(A)$ 大小的限制，极限问题

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^+$$

总是可以考虑的. 设

$$A=\begin{bmatrix}1 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix},$$

则 $\operatorname{rank}(A)=1$，它是亏秩的. 容易算出，

$$A^+=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

如果令

$$\delta A=\begin{bmatrix}0 & 0\\ 0 & \varepsilon\\ 0 & 0\end{bmatrix},\quad \varepsilon\neq 0,$$

则

$$A+\delta A=\begin{bmatrix}1 & 0\\ 0 & \varepsilon\\ 0 & 0\end{bmatrix},$$

而矩阵 $A+\delta A$ 为列满秩，即 $\operatorname{rank}(A+\delta A)=2>\operatorname{rank}(A)$，可以算出

$$(A+\delta A)^+=\begin{bmatrix}1 & 0 & 0\\ 0 & \varepsilon^{-1} & 0\end{bmatrix}.$$

从而可知当 $\varepsilon\to 0$ 时(这时 $\|\delta A\|\to 0$)，极限

$$\lim_{\varepsilon\to 0}(A+\delta A)^+$$

根本就不存在. 但是，如果我们令

$$\delta A=\begin{bmatrix}1 & \varepsilon\\ 0 & 0\\ 0 & 0\end{bmatrix},\quad \varepsilon\neq 0,$$

则 $\operatorname{rank}(A+\delta A)=\operatorname{rank}(A)=1$，容易算出

$$(A+\delta A)^+=\frac{1}{1+\varepsilon^2}\begin{bmatrix}1 & 0 & 0\\ \varepsilon & 0 & 0\end{bmatrix}.$$

在这种情形，当 $\varepsilon\to 0$ 时却有

$$\lim_{\varepsilon\to 0}(A+\delta A)^+=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}=A^+.$$

但总的说来，例中的矩阵 A，它的广义逆矩阵 A^+ 是不连续的.

3-2　满秩矩阵广义逆的连续性

上一小节中的例虽然是一个特例，它却反映了情况：第一种情形，$\operatorname{rank}(A+\delta A)>\operatorname{rank}(A)$，这时极限不存在，而第二种情形，$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A)$，这时极限存在而且就是 A^+. 这两种情形的对比是很有意义的，它说明极限存在与否，似乎与矩阵秩是否发生变化有关. 对于满秩矩阵来说，只要摄动量足够小，矩阵总恒保持为满秩，方阵的情形就说明了这个事实. 实际上，和满秩方阵的情形一样，满秩的 $m\times n$ 阶矩阵 A，它的广义逆矩阵 A^+ 的确也是连续的. 我们有

定理 6　设 A 是一个满秩的 $m\times n$ 阶矩阵，则恒有

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^+=A^+. \tag{31}$$

证明　不失一般性可设 A 为列满秩，即 $\operatorname{rank}(A)=n$. 所以，A^TA 为 $n\times n$ 阶非奇异矩阵（实际上为对称正定）. 这时 A^+ 可以表示成

$$A^+=(A^TA)^{-1}A^T.$$

记 δA 为同阶摄动矩阵，我们有关系

$$(A+\delta A)^T(A+\delta A)=A^TA+(A+\delta A)^T\delta A+(\delta A)^TA.$$

根据 Banach 引理的推论可知，如果

$$\|(A^TA)^{-1}[(A+\delta A)^T(\delta A)+(\delta A)^TA]\|<1,$$

则矩阵$(A+\delta A)^T(A+\delta A)$为非奇异方阵. 另一方面，我们总可以取$\|\delta A\|$足够小，以保证上述不等式成立. 因为 A 是给定的，存在有正数 η，只要$\|\delta A\|\leqslant\eta$，上述不等式总成立. 这时，广义逆矩阵$(A+\delta A)^+$ 可以表示成

$$(A+\delta A)^+=[(A+\delta A)^T(A+\delta A)]^{-1}(A+\delta A)^T.$$

令 $\|\delta A\|\to 0$，这时，一方面有

$$\lim_{\|\delta A\|\to 0}[(A+\delta A)^T(A+\delta A)]^{-1}=(A^TA)^{-1},$$

同时还有

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^T=A^T,$$

从而便证明

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^+=(A^TA)^{-1}A^T=A^+.$$

这就是本定理的结论. 证毕.

§4　亏秩矩阵广义逆的连续性问题

4-1　摄动不改变矩阵秩的情形

在前一节的例中已表现出摄动对矩阵秩的影响和极限

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^+$$

存在与否有一定的联系. 在这一节中，我们将严格证明前面的预测是完全正确的. 在这一小节中我们将先考虑摄动 δA 不改变 $A+\delta A$ 的秩的情形，即 δA 恒保持

$$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A)=k$$

的情形. 不失一般性，可假定 $m\geqslant n$，这时

$$k<n.$$

我们首先对$\|(A+\delta A)^+\|$进行估计.

引理 1 设

$$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A)=k<n, \tag{a}$$

且

$$\|A^+\|\|\delta A\|<1① \tag{b}$$

我们有估计式:

$$\|(A+\delta A)^+\|\leqslant\frac{\|A^+\|}{1-\|A^+\|\|\delta A\|}. \tag{32}$$

证明 因为 A 和 $A+\delta A$ 同秩,所以它们有同样多的正奇异值,分别记之为

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_k \quad 和 \quad \sigma_1\geqslant\sigma_2\geqslant\cdots\geqslant\sigma_k.$$

根据§1中的奇异值摄动定理可知

$$\sigma_k\geqslant\lambda_k-\|\delta A\|.$$

另外,我们已知

$$\|A^+\|=\frac{1}{\lambda_k},\quad \|(A+\delta A)^+\|=\frac{1}{\sigma_k},$$

这就是说,

$$\lambda_k=\frac{1}{\|A^+\|},\quad \sigma_k=\frac{1}{\|(A+\delta A)^+\|}.$$

从而便得到不等式

$$\frac{1}{\|(A+\delta A)^+\|}\geqslant\frac{1}{\|A^+\|}-\|\delta A\|.$$

因为$\|A^+\|\|\delta A\|<1$,于是就得到估计式

$$\|(A+\delta A)^+\|\leqslant\frac{\|A^+\|}{1-\|A^+\|\|\delta A\|}. \quad 证毕.$$

引理 2 记 $B=A+\delta A$,我们有以下的一般关系式:

---

① 在这个条件下,可以证明恒有

$$\operatorname{rank}(A+\delta A)\geqslant\operatorname{rank}(A).$$

$$B^+ - A^+ = -B^+(\delta A)A^+ + B^+B^{+T}(\delta A)^T(I_m - AA^+) + (I_n - B^+B)(\delta A)^T A^{+T} A^+. \tag{33}$$

证明 首先我们把 $B^+ - A^+$ 表示成:

$$B^+ - A^+ = -B^+(B-A)A^+ + (B^+ - A^+) + B^+(B-A)A^+.$$

把这个恒等式的右端依次重新组合各项如下:

$$\begin{aligned} &-B^+(\delta A)A^+ + B^+ - B^+AA^+ + B^+BA^+ - A^+ \\ &= -B^+(\delta A)A^+ + B^+(I_m - AA^+) - (I_n - B^+B)A^+ \\ &= B^+ - A^+, \end{aligned} \tag{34}$$

其中, I_m 和 I_n 分别表示 $m\times m$ 阶和 $n\times n$ 阶单位矩阵.

根据第 2 章 **5-3** 中关于 A^+ 的性质 6°,

$$A^T(I_m - AA^+) = A^T - A^TAA^+ = A^T - A^T = 0.$$

由此可推得

$$\begin{aligned} B^+(I_m - AA^+) &= B^+BB^+(I_m - AA^+) \\ &= B^+(BB^+)^T(I_m - AA^+) \\ &= B^+B^{+T}B^T(I_m - AA^+) \\ &= B^+B^{+T}(A^T + \delta A^T)(I_m - AA^+) \\ &= B^+B^{+T}(\delta A)^T(I_m - AA^+). \end{aligned} \tag{35}$$

用同样的方法可以证明

$$-(I_n - B^+B)A^+ = (I_n - B^+B)(\delta A)^T A^{+T} A^+. \tag{36}$$

把(35), (36)代入(34)中, 就得到了所需要的恒等关系(33). 证毕.

特例 当矩阵满秩时, 引理中的恒等式还可简化. 例如, 若 $\mathrm{rank}(A) = m$, 则 $AA^+ = I_m$, 于是(33)就简化为

$$B^+ - A^+ = -B^+(\delta A)A^+ + (I_n - B^+B)(\delta A)^T A^{+T} A^+. \tag{c}$$

其中, 若 A、B 都是 $n\times n$ 阶非奇异方阵, 则(c)就变成

$$B^{-1} - A^{-1} = -B^{-1}(\delta A)A^{-1}. \tag{d}$$

有了上述引理，我们便可以证明

定理7 对一般的 $m\times n$ 阶矩阵 A 和 B，总有估计式：

$$\|B^+-A^+\|\leqslant(\|A^+\|\|B^+\|+\|B^+\|^2+\|A^+\|^2)\|B-A\|. \quad (37)$$

证明 设 A、B 均为亏秩，则 $AA^+\neq I_m$，$B^+B\neq I_n$，这时，I_m-AA^+ 和 I_n-B^+B 均为直交投影矩阵，它们的范数都是1. 对恒等式(33)两端取范数就得到(37). 证毕.

如果 $AA^+=I_m$ 或 $B^+B=I_n$，则(37)亦然成立，而且结果要更好一些. 例如，若 $AA^+=I_m$，则(33)变成

$$\begin{aligned}B^+-A^+=&-B^+(\delta A)A^+\\&+(I_n-B^+B)(\delta A)^TA^{+T}A^+,\end{aligned} \quad (38\text{-}1)$$

于是

$$\begin{aligned}\|B^+-A^+\|&\leqslant\|B^+\|\|A^+\|\|\delta A\|+\|\delta A\|\cdot\|A^+\|^2\\&=(\|B^+\|\|A^+\|+\|A^+\|^2)\|B-A\|.\end{aligned}$$

若 $B^+B=I_n$，则有

$$\|B^+-A^+\|\leqslant(\|B^+\|\|A^+\|+\|B^+\|^2)\|B-A\|. \quad (38\text{-}2)$$

现在我们来考虑亏秩矩阵广义逆的连续性问题.

首先，从估计式(32)可知，若 $\|A^+\|\|\delta A\|<1$，且

$$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A),$$

则 $\|(A+\delta A)^+\|$ 有界. 所以，当 $\|\delta A\|\to 0$ 时，恒有

$$\lim_{\|\delta A\|\to 0}\|(A+\delta A)^+-A^+\|=0,$$

即

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^+=A^+.$$

应当指出，在取极限 $\|\delta A\|\to 0$ 时，δA 必须恒保持矩阵 $A+\delta A$ 的秩和 A 的秩相等.

上述结论可以表示成下述定理的形式.

定理8 设 $\{A_i\}$ 是一个 $m\times n$ 阶矩阵序列，其极限为矩阵 A，即

$$\lim_{i\to\infty}A_i=A. \quad (39)$$

欲

$$\lim_{i\to\infty} A_i^+ = A^+, \tag{40}$$

它的一个充分条件是：对充分大的 i，例如当 $i>N$（某一个给定的自然数）时，恒有

$$\operatorname{rank}(A_i) = \operatorname{rank}(A). \tag{41}$$

4-2 摄动使亏秩矩阵的秩增大的情形

如果 $m\times n$ 阶矩阵 A 为亏秩，$\operatorname{rank}(A)=k$，当摄动 δA 充分小时，$A+\delta A$ 的秩不会减小。这是因为，若记 $A=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$，并假定前 k 列为线性独立，令 $\delta A=[\delta\boldsymbol{a}_1, \cdots, \delta\boldsymbol{a}_n]$，则当 $\max\limits_i\{\|\delta\boldsymbol{a}_i\|\}$ 充分小时，容易证明向量系 $\boldsymbol{a}_1+\delta\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k+\delta\boldsymbol{a}_k$ 恒保持线性独立，这就证明，当 $\|\delta A\|$ 充分小时，$\operatorname{rank}(A+\delta A)\geqslant k$。因此，对亏秩矩阵我们只考虑摄动使矩阵的秩增大时会产生什么样的影响。为此目的，我们假定

$$\operatorname{rank}(A+\delta A) > \operatorname{rank}(A) = k. \tag{42}$$

在这种情况下我们有以下的估计式。

定理 9 若亏秩矩阵 A 的摄动矩阵 $A+\delta A$ 满足条件 (42)，则必有估计式：

$$\|(A+\delta A)^+\| \geqslant \frac{1}{\|\delta A\|}. \tag{43}$$

证明 设 $\operatorname{rank}(A+\delta A)=r>k$，则矩阵 A 的第 r 个奇异值 $\lambda_r=0$。这时，根据奇异值摄动定理，$A+\delta A$ 的第 r 个奇异值（即最小的正奇异值）σ_r 满足不等式

$$\sigma_r \leqslant \|\delta A\|. \tag{44}$$

因为

$$\|(A+\delta A)^+\| = \frac{1}{\sigma_r},$$

所以有

$$\|(A+\delta A)^+\| \geqslant \frac{1}{\|\delta A\|},$$

即 (43)。证毕。

这个定理说明，如果摄动 δA 使矩阵 $A+\delta A$ 的秩增大，$\|\delta A\|$ 愈小，$\|(A+\delta A)^+\|$ 就愈大，当 $\|\delta A\|\to 0$ 时，$\|(A+\delta A)^+\|\to\infty$.

推论 设 $\{A_i\}$ 是一个 $m\times n$ 阶矩阵序列，且

$$\lim_{i\to\infty} A_i = A,$$

欲

$$\lim_{i\to\infty} A_i^+ = A^+,$$

其必要条件为：对充分大的 i，恒有

$$\operatorname{rank}(A_i)=\operatorname{rank}(A).$$

综合以前的讨论，我们就得到了

定理 18(Stewart) 设 A 为任意的 $m\times n$ 阶矩阵，欲

$$\lim_{\|\delta A\|\to 0} (A+\delta A)^+ = A^+,$$

其充分必要条件为：对充分小的 $\|\delta A\|$，恒有

$$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A).$$

前面我们已经讲过，对于充分小的 δA，$\operatorname{rank}(A+\delta A)$ 不可能小于 $\operatorname{rank}(A)$. 现在我们来进一步证明，对任意给定的正数 η，不论它如何小，如果矩阵 A 为亏秩，则总存在摄动 δA，它一方面满足不等式

$$\|\delta A\|\leqslant\eta,$$

而同时又增大了矩阵 $A+\delta A$ 的秩，即

$$\operatorname{rank}(A+\delta A)>\operatorname{rank}(A).$$

今证明于下.

记 $A=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$，并设 $\operatorname{rank}(A)=k<n$（这里仍假定 $m\geqslant n$），不失一般性，可假定 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为线性独立，并且 $\boldsymbol{e}_i$ 不属于 $\operatorname{Span}\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k\}$. 令

$$\delta A=[\underbrace{\mathbf{0}, \cdots, \mathbf{0}}_{k\text{个}}, \ \varepsilon\boldsymbol{e}_i, \ \underbrace{\mathbf{0}, \cdots, \mathbf{0}}_{n-k-1\text{个}}], \quad \varepsilon\neq 0.$$

容易证明 $\|\delta A\|=|\varepsilon|$，而不论 $|\varepsilon|$ 怎样小，恒有

$$\operatorname{rank}(A+\delta A)=k+1>\operatorname{rank}(A).$$

于是我们便证明了一个重要结论.

定理 11 亏秩矩阵的广义逆是不连续的.

§5 亏秩矩阵的保秩变形和广义逆矩阵不连续性的消除

5-1 亏秩矩阵的保秩变形

从上面的讨论可以看出亏秩矩阵广义逆矩阵的不连续性，其根源在于微小摄动就可以增大矩阵的秩. 我们还看到这种秩的增大似乎又是不可避免的. 如此说来，在电子计算机上计算亏秩矩阵的广义逆矩阵是不可能的了. 我们指出，这种纯理论的结论，实际上是似是而非的. 换句话说，亏秩矩阵广义逆的这种不连续性，仅仅是一种迷糊人的表面现象，并非实质性的. 经过计算数学上的适当处理，由于矩阵亏秩而产生的不连续性，完全可以消除. 为了消除这种所谓的不连续性，我们将提出一种亏秩矩阵的保秩变形的思想，以及实现这种保秩变形的可行的方法.

亏秩矩阵的保秩变形 今假定 A 为给定的一个 $m\times n$ 阶亏秩矩阵，$\operatorname{rank}(A)=k$. 其次设 δA 为给定的任意的微小摄动量. 现在我们根据 $A+\delta A$ 来构造一个同阶的矩阵 A_δ，使之具有下列性质：

(i) A_δ 是对 $A+\delta A$ 加以适当处理而得到的，不是独立于 $A+\delta A$ 之外而随意臆造的.

(ii) 当 $\|\delta A\|\to 0$ 时，$A_\delta\to A$.

(iii) 当 $\|\delta A\|$ 充分小时，恒有 $\operatorname{rank}(A_\delta)=\operatorname{rank}(A)$.

定义 3 具有上列三种性质的矩阵 A_δ 为由矩阵 A 和摄动 δA 所确定的(不是唯一的)矩阵 $A+\delta A$ 的一种**保秩变形**. 说得更确切一些，我们可以称 A_δ 为 $A+\delta A$ 的一种**有效的保秩变形**.

在给出实现矩阵的保秩变形的具体方法之前，有必要对这个定义作一些补充说明. 关于定义中的(i)，我们特别强调 A_δ 是经过对 $A+\delta A$ 适当加以处理而得到的，不是独立于 $A+\delta A$ 之外而随意臆造的. 比如说，当 δA 给定以后，我们可以简单地规定 $A_\delta=A+\|\delta A\|A$，它显然满足定义中的性质(ii)和(iii)，然而，这样构造的保秩变形矩阵对于广义逆矩阵的计算来说，毫无意义，因而是一种无效的保秩变形. 我们之所以要建立保秩变形矩阵这个概念，是有明确的目的，即为实际计算亏秩矩阵的广义逆及与其相关联的线性最小二乘问题等实际计算问题服务的. 在定义的最后，我们所以特别加上"有效的"这三个字，并非多余的，正是为了强调这里所指出的目的性. 但是，为了行文简练，我们就用"保秩变形"这个词，而略去"有效的"三字，它所包含的意义不变!

5-2 保秩变形的实现方法

考虑到今后的实际需要，从计算数学的观点出发，在这一小节中我们将给出构造保秩变形矩阵的具体方法.

构造保秩变形矩阵的直交化方法 我们先把给定的亏秩矩阵($m\times n$ 阶)A 用它的 n 个 m 维列向量表示成

$$A=[\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_n],$$

我们假定:

1° $\mathrm{rank}(A)=k$.

2° 不失一般性，我们假定 $m\geqslant n$，并且前 k 列

$$\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_k$$

为线性独立.

用 Gram-Schmidt 直交化方法把向量序列 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 化为一个标准直交系 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k$. 其中的每一个向量 $\boldsymbol{q}_j$ 都是前 j 个向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_j$ 的线性组合, 并和前 $j-1$ 个列向量直交

$$\boldsymbol{q}_j^T \boldsymbol{a}_i = 0, \ i=1, 2, \cdots, j-1.$$

构造 $m \times k$ 阶满秩矩阵 Q_k,

$$Q_k = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k], \tag{45}$$

则 $Q_k^T Q_k = I_k$. 用 Q_k^T 左乘矩阵 A 便得到

$$Q_k^T A = [R \vdots U], \tag{46}$$

其中的 R 是一个 $k \times k$ 阶的非奇异上三角阵, 而 U 是一个一般形式的 $k \times (n-k)$ 阶矩阵. 显然有

$$\mathrm{rank}(Q_k^T A) = \mathrm{rank}([R \vdots U]) = k. \tag{47}$$

在另一方面, 不难证明:

$$A = Q_k [R \vdots U]. \tag{48}$$

把摄动矩阵表示成:

$$A + \delta A = [\boldsymbol{a}_1 + \delta \boldsymbol{a}_1, \cdots, \boldsymbol{a}_n + \delta \boldsymbol{a}_n]. \tag{49}$$

仿前, 把 $\boldsymbol{a}_1 + \delta \boldsymbol{a}_1, \cdots, \boldsymbol{a}_k + \delta \boldsymbol{a}_k$ 用 Gram-Schmidt 方法化为一个标准直交系 $\tilde{\boldsymbol{q}}_1, \cdots, \tilde{\boldsymbol{q}}_k$, 然后令

$$Q_\delta = [\tilde{\boldsymbol{q}}_1, \cdots, \tilde{\boldsymbol{q}}_k], \tag{50}$$

用 Q_δ^T 左乘 $A + \delta A$ 得

$$Q_\delta^T [A + \delta A] = [\tilde{R} \vdots \tilde{U}], \tag{51}$$

其中, $\tilde{R}$ 是一个 $k \times k$ 阶上三角阵, $\tilde{U}$ 是 $k \times (n-k)$ 阶矩阵. 显然, 当 $\|\delta A\| \to 0$ 时, $\|\delta \boldsymbol{a}_i\| \to 0$, $i=1, 2, \cdots, n$, 从而

$$\tilde{\boldsymbol{q}}_i \to \boldsymbol{q}_i, \ i=1, 2, \cdots, k.$$

这同时也表示 $Q_\delta \to Q_k$. 又

$$[\tilde{R} \vdots \tilde{U}] - [R \vdots U] = Q_\delta^T (A + \delta A) - Q_k^T A,$$

所以, 当 $\|\delta A\| \to 0$, 也必有 $[\tilde{R} \vdots \tilde{U}] \to [R \vdots U]$, 从而可知, 当

$\|\delta A\|$ 充分小时，$\widetilde{R}$ 为非奇异．令

$$A_\delta = Q_r[\widetilde{R} : \widetilde{U}], \tag{52}$$

则当 $\|\delta A\|$ 充分小时，恒有

$$\operatorname{rank}(A_\delta) = \operatorname{rank}(A) = k, \tag{53}$$

并且，当 $\|\delta A\| \to 0$ 时

$$\lim_{\|\delta A\| \to 0} A_\delta = A. \tag{54}$$

这样一来，我们便得到了满足定义中条件(ii)和(iii)的保秩变形矩阵 A_δ．根据 Stewart 定理可知，

$$\lim_{\|\delta A\| \to 0} A_\delta^+ = A^+.$$

这就是说，当 $\|\delta A\|$ 足够小时，A_δ^+ 便是 A^+ 的一种较好的近似。

我们说这样构造出来的 A_δ 是矩阵 $A+\delta A$ 的一种有效的保秩变形．因为 A_δ 是利用直交化方法对矩阵 $A+\delta A$ 进行矩阵分解的过程中确定的，而直交化方法又是计算广义逆矩阵及与其有关联的线性最小二乘问题一种重要方法．

在这里虽然我们只举出了一种构造保秩变形矩阵的方法，但是我们相信，细心的读者会由此得到启发，收到举一反三的效果．比如说，直交化方法从某种意义上讲，不过是实现矩阵的满秩分解的手段之一，而任何实现满秩分解的方法，原则上都可以用来构造亏秩矩阵的保秩变形．

最后，我们再指出一点，目前用来计算广义逆矩阵以及线性最小二乘问题的一切直接方法，除正则化方法外，都隐含着保秩变形的思想，从而消除了摄动增秩所带来的严重危害，在良态的情况下，可以收到比较好的效果．当然，不同的算法所取得的实际效果也会有很大差别，但这种情况的发生，是由于各种算法本身的数值稳定性程度有所差异而引起的．这是另

外一种性质的问题.

§6 最小二乘问题的摄动定理

在本章最后一节，我们将就摄动对最小二乘问题解的影响作出估计.

今假定

$$Ax=\boldsymbol{b} \tag{55}$$

为给定的超定线性方程组，系数矩阵 A 为 $m\times n$ 阶，$\boldsymbol{b}$ 为 m 维已知向量，$\boldsymbol{x}$ 为 n 维待定向量. 这个超定方程组的极小最小二乘解用 A 的广义逆矩阵 A^+ 可以表示成：

$$\boldsymbol{x}=A^+\boldsymbol{b}. \tag{56}$$

如果系数矩阵以及右端向量 $\boldsymbol{b}$ 分别有摄动 δA 和 $\delta\boldsymbol{b}$，相应摄动方程

$$(A+\delta A)(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b}+\delta\boldsymbol{b} \tag{57}$$

的极小最小二乘解为

$$\boldsymbol{x}+\delta\boldsymbol{x}=(A+\delta A)^+(\boldsymbol{b}+\delta\boldsymbol{b}). \tag{58}$$

假定摄动 δA 不改变矩阵的秩，并且 $\|\delta A\|$ 适当地小，我们来对由于摄动对解所产生的影响 $\delta\boldsymbol{x}$ 进行估计. 对此有

定理 12 假定下列条件成立：

1° $\operatorname{rank}(A+\delta A)=\operatorname{rank}(A)$.

2° $\|A^+\|\|\delta A\|<1$.

则有估计式：

$$\|\delta\boldsymbol{x}\|\leqslant\frac{\|A^+\|}{1-\Delta}\left(2\|\delta A\|\|\boldsymbol{x}\|+\|\delta\boldsymbol{b}\|+\frac{\Delta\|\boldsymbol{r}\|}{1-\Delta}\right), \tag{59}$$

其中,

$$\Delta=\|A^+\|\|\delta A\|,\ \boldsymbol{r}=\boldsymbol{b}-A\boldsymbol{x}. \tag{60}$$

证明 从公式(58)的两端分别减去(56)的左、右端，便得

$$\delta \boldsymbol{x}=\{(A+\delta A)^{+}-A^{+}\}\boldsymbol{b}+(A+\delta A)^{+}\delta\boldsymbol{b}.$$

根据 **4-2** 的引理 2 中的公式(33)，上式可以表示成：

$$\delta \boldsymbol{x}=-B^{+}(\delta A)\boldsymbol{x}+B^{+}B^{+T}(\delta A)^{T}\boldsymbol{r}$$
$$+(I-B^{+}B)(\delta A)^{T}A^{+T}\boldsymbol{x}+B^{+}(\delta\boldsymbol{b}), \qquad (61)$$

其中，$B=A+\delta A$.

据条件 2°，$\|A^{+}\|\|\delta A\|<1$，所以 $\|B^{+}\|$ 有界：

$$\|B^{+}\|\leqslant\frac{\|A^{+}\|}{1-\|A^{+}\|\|\delta A\|}.$$

（参看 **4-2** 引理 1.）

对(61)两端分别取范数得：

$$\|\delta\boldsymbol{x}\|\leqslant\|B^{+}\|\|\delta A\|\|\boldsymbol{x}\|+\|B^{+}\|^{2}\|\delta A\|\|\boldsymbol{r}\|$$
$$+\|\delta A\|\|A^{+}\|\|\boldsymbol{x}\|+\|B^{+}\|\|\delta\boldsymbol{b}\|$$
$$=\|B^{+}\|(\|\delta A\|\|\boldsymbol{x}\|+\|\delta\boldsymbol{b}\|)+\|\delta A\|\|A^{+}\|\|\boldsymbol{x}\|$$
$$+\|B^{+}\|^{2}\|\delta A\|\|\boldsymbol{r}\|$$
$$\leqslant\frac{\|A^{+}\|}{1-\Delta}(\|\delta A\|\|\boldsymbol{x}\|+\|\delta\boldsymbol{b}\|)+\|A^{+}\|\|\delta A\|\|\boldsymbol{x}\|$$
$$+\frac{\|A^{+}\|^{2}}{(1-\Delta)^{2}}\|\delta A\|\|\boldsymbol{r}\|$$
$$=\frac{\Delta\|\boldsymbol{x}\|+\|A^{+}\|\|\delta\boldsymbol{b}\|}{1-\Delta}+\frac{\Delta\|A^{+}\|\|\boldsymbol{r}\|}{(1-\Delta)^{2}}+\Delta\|\boldsymbol{x}\|$$
$$=\frac{\Delta\|\boldsymbol{x}\|+\|A^{+}\|\|\delta\boldsymbol{b}\|+\Delta\|\boldsymbol{x}\|-\Delta^{2}\|\boldsymbol{x}\|}{1-\Delta}$$
$$+\frac{\Delta\|A^{+}\|\|\boldsymbol{r}\|}{(1-\Delta)^{2}}$$
$$=\frac{1}{1-\Delta}\Big(2\Delta\|\boldsymbol{x}\|+\|A^{+}\|\|\delta\boldsymbol{b}\|-\Delta^{2}\|\boldsymbol{x}\|$$
$$+\frac{\Delta\|A^{+}\|\|\boldsymbol{r}\|}{1-\Delta}\Big)$$

把末端分子中的 Δ 再换成 $\|A^{+}\|\|\delta A\|$，就得到估计式：

$$\|\delta \boldsymbol{x}\| \leqslant \frac{\|A^+\|}{1-\Delta}\Big(2\|\delta A\|\|\boldsymbol{x}\| - \|A^+\|\|\delta A\|^2\|\boldsymbol{x}\|$$

$$+ \|\delta \boldsymbol{b}\| + \frac{\Delta\|\boldsymbol{r}\|}{1-\Delta}\Big)$$

$$\leqslant \frac{\|A^+\|}{1-\Delta}\Big(2\|\delta A\|\|\boldsymbol{x}\| + \|\delta \boldsymbol{b}\| + \frac{\Delta\|\boldsymbol{r}\|}{1-\Delta}\Big).$$

于是便证明了估计式(59). 证毕.

从这个估计式可以看出,如果残量 $\|\boldsymbol{r}\| = \|\boldsymbol{b} - A\boldsymbol{x}\|$ 很大,则摄动对解的影响也就越大.

推论 把 $\boldsymbol{b}$ 分解成 $\boldsymbol{b} = P_{\mathscr{R}(A)}\boldsymbol{b} + P_{\mathscr{R}^\perp(A)}\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2$. 根据 $A\boldsymbol{x} = \boldsymbol{b}_1$, 从而有 $\|\boldsymbol{b}_1\| = \|A\boldsymbol{x}\| \leqslant \|A\|\|\boldsymbol{x}\|$, 即

$$\frac{1}{\|\boldsymbol{x}\|} \leqslant \frac{\|A\|}{\|\boldsymbol{b}_1\|}.$$

乘(59)两端便得到相对摄动的估计式:

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqslant \frac{k(A)}{1-\Delta}\Big[\Big(2 + \frac{k(A)}{1-\Delta}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}_1\|}\Big)\frac{\|\delta A\|}{\|A\|} + \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}_1\|}\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}\Big]. \tag{62}$$

其中, $k(A) = \|A\|\|A^+\|$.

用同样的方法,不难证明下列结论:

(i) 超定满秩情形:即 $m > n$, $\operatorname{rank}(A) = n$.

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqslant \frac{k(A)}{1-\Delta}\Big[\Big(1 + \frac{k(A)}{1-\Delta}\frac{\|\boldsymbol{r}\|}{\|\boldsymbol{b}_1\|}\Big)\frac{\|\delta A\|}{\|A\|} + \frac{\|\boldsymbol{b}\|}{\|\boldsymbol{b}_1\|}\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}\Big]. \tag{63}$$

(ii) 常定满秩的情形, 即 $m = n = \operatorname{rank}(A)$.

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqslant \frac{k(A)}{1-\Delta}\Big(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}\Big). \tag{64}$$

(iii) 亚定行满秩的情形, 即 $m < n$, $\operatorname{rank}(A) = m$.

$$\frac{\|\delta \boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\frac{k(A)}{1-\Delta}\left(2\frac{\|\delta A\|}{\|A\|}+\frac{\|\delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}\right). \tag{65}$$

从以上这些估计式可以看出,(ii),(iii)两种情形,摄动的影响比较小,而这两种情形相应的方程组都相容.在(i)中,如果方程组相容,则$\boldsymbol{r}=0$, $\boldsymbol{b}_1=\boldsymbol{b}$,于是估计式(63)就化为(64).

最后,我们要着重指出,上述对摄动的估计是针对原始的最小二乘问题而言的.在实际进行计算时,原来的问题已经发生了变化.因此,摄动的影响应对新的计算问题来考虑.就我们在第5章中所介绍的算法而言,最后总是把问题归结为相容的线性方程组求解问题,从而便消除了残量的影响.

第 4 章

病态问题和病态程度的度量

§1 病态问题和算法的数值稳定性

1-1 计算问题的病态和良态

一个计算问题，例如关于线性方程组

$$A\boldsymbol{x}=\boldsymbol{b} \quad (\boldsymbol{x},\boldsymbol{b}\in\mathscr{R}^n,\ A\text{ 为 } n\times n \text{ 阶}) \tag{1}$$

的求解问题，或者与其相关联的矩阵 A 的求逆问题，如果矩阵 A 为非奇异，则逆矩阵 A^{-1} 在理论上是存在的，方程组(1)的解 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$ 是唯一的．但是，若系数矩阵 A 有摄动 δA，右端向量 $\boldsymbol{b}$ 有摄动 $\delta\boldsymbol{b}$，那么，相应的摄动方程组的解也和 $\boldsymbol{x}=A^{-1}\boldsymbol{b}$ 不同．设其解为 $\boldsymbol{x}+\delta\boldsymbol{x}$，即 $\boldsymbol{x}+\delta\boldsymbol{x}$ 准确地满足摄动方程

$$(A+\delta A)(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b}+\delta\boldsymbol{b}. \tag{2}$$

根据 Banach 引理的推论可知，从理论上讲，只要 $\|A^{-1}\|\|\delta A\|<1$，则摄动矩阵 $A+\delta A$ 总是非奇异的，于是，$\boldsymbol{x}+\delta\boldsymbol{x}$ 可以表示成

$$\boldsymbol{x}+\delta\boldsymbol{x}=(A+\delta A)^{-1}(\boldsymbol{b}+\delta\boldsymbol{b}). \tag{3}$$

我们已经知道，非奇异方阵的逆矩阵是连续的，这就是说，

$$\lim_{\|\delta A\|\to 0}(A+\delta A)^{-1}=A^{-1},$$

从而，当 $\|\delta b\|$，$\|\delta A\|$ 都趋于零时，

$$\lim_{\substack{\|\delta A\|\to 0\\ \|\delta \boldsymbol{b}\|\to 0}}(\boldsymbol{x}+\delta\boldsymbol{x})=\lim_{\substack{\|\delta A\|\to 0\\ \|\delta \boldsymbol{b}\|\to 0}}(A+\delta A)^{-1}(\boldsymbol{b}+\delta\boldsymbol{b})$$

$$=A^{-1}\boldsymbol{b}=\boldsymbol{x}.$$

这就是说，只要 $\|\delta A\|$ 和 $\|\delta\boldsymbol{b}\|$ 足够小，总可以使解的摄动量

$\|\delta x\|$ 小到任意规定的程度，而逆矩阵 $(A+\delta A)^{-1}$ 也可以逼近 A^{-1} 到任意的精确程度. 实际上这仅仅是一种纯理论的结果. 从实际计算的角度来看，问题远远不是这样简单. 为了进一步弄清楚这个问题，我们先指出以下两点：

1° 摄动是不可避免的. 其原因有二. 第一，问题中的数据往往是由实验观测（对最小二乘问题尤其如此）得到的，或者是通过计算得到的，无论是那一种情形，都会出现观测误差或计算误差. 第二，我们所使用的通用电子计算机的字长总是有限的，数据输入到计算机中以及在运算的过程中，舍入误差也是不可避免的，原始数据即使完全精确，在计算机中也不一定能丝毫没有误差的表示出来. 这两种情况所产生的影响，相当于原始数据的摄动. 我们计算得到的结果，实际上是某一摄动问题的精确解.

2° 单从计算机字长有限这一现实来看，摄动量的相对精度，例如 $\|\delta A\|/\|A\|$ 或 $\|\delta \boldsymbol{b}\|/\|\boldsymbol{b}\|$，是有一定的局限的，它们并不能任意地小.

根据上述事实，对于一个给定的计算问题，我们必须研究的一个重要问题是：问题中参数的微小摄动，对问题的解会产生什么样的影响. 这就是问题的（理论）解对于参数摄动的敏感性问题. 为了具体地说明这个问题，今举一例.

例 (R. S. Wilson) 今考虑系数矩阵为 4×4 阶实对称矩阵的线性方程组：

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix}, \tag{4}$$

它的解为 $[x_1, x_2, x_3, x_4] = [1, 1, 1, 1]$.

如果把右端项作微小摄动，

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1+\delta x_1 \\ x_2+\delta x_2 \\ x_3+\delta x_3 \\ x_4+\delta x_4 \end{bmatrix} = \begin{bmatrix} 32.1 \\ 22.9 \\ 33.1 \\ 30.9 \end{bmatrix}, \tag{5}$$

它的解就变为

$$[9.2,\ -12.6,\ 4.5,\ -1.1]^T.$$

从这个例子可以看出，右端仅有约1/200的相对摄动，(即用摄动后的右端作为原右端项的近似时的相对误差)，而解的相对误差却放大了约2000倍！这就说明线性方程组(4)的解对参数的微小摄动十分敏感. 无论如何我们也不会把(5)的解当成(4)的近似解.

如果把(4)的系数微加摄动：

$$\begin{bmatrix} 10 & 7 & 8.1 & 7.2 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{bmatrix} \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \\ \tilde{x}_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix}, \tag{6}$$

其解为 $[-81\ \ 137\ \ -34\ \ 22]$,

这和(4)的解的差异就更大了.

有一类矩阵叫做 Hilbert 矩阵，其定义为：

$$H_n = \begin{bmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n+1} \\ \frac{1}{n} & \frac{1}{n+1} & \cdots & \frac{1}{2n-1} \end{bmatrix}. \tag{7}$$

以 H_n 为系数矩阵的方程组，或以其前 k 列组成的 $n\times k$ 阶矩

阵作为系数矩阵的线性超定方程组，也具有和线性方程组(4)类似的性质，阶数越高，微小摄动对解的影响也就越为严重．

值得指出的是，(4)和(7)中的矩阵不但是对称的，而且还都是正定的呢！

上例说明，的确存在有这样的计算问题，其中的参数有微小的相对摄动时，解就会引起"巨大"的相对摄动．这种情形，我们称之为问题本身的病态性质．反之，如果问题中参数的微小相对摄动，对解所引起的相对摄动不"大"，便称这种计算问题为良态．

这里我们还要着重说明以下几点：

1）一个数学计算问题的病态或良态，是这个数学问题本身所固有的一种属性，问题的病态对它的计算会带来困难，但病态性这个概念本身并不直接涉及计算这个问题的算法．在我们打算对某一数学问题进行实际计算的时候，对它的病态性是必须加以考虑的．

2）病态和良态之间并没有明显的界限，上面所讲的病态或良态的概念仅仅是一种"定性的"描述，从实际计算的角度来看，我们还必须设法从"定量"上加以考虑．这就是说，我们得设法对参数的一定的相对摄动，对解的相对摄动作出估计，从而判断问题的病态程度．

3）病态问题是计算数学的各领域中普遍存在的问题，例如，代数方程求根，微分方程求数值解，以及插值和数值积分等计算问题，都有所谓病态问题，因而研究问题的病态性质以及它们的处理方法，是计算数学中一类十分重要的课题．在本章中我们将只对本书所涉及的线性问题进行初步的探讨．

1-2 算法的数值稳定性概念

在对问题的病态程度进行估计之前，我们先来对计算一

个数学问题的"算法"以及与其有关的"数值稳定性"概念作简单的介绍. 所谓计算一个数学问题的算法，概略地说，它是某一计算方法的具体实现. 为了说明这个概念，今举一例.

例　二次方程求根问题. 今给定一个二次实系数代数方程 $x^2+px+q=0$. 它的两个根可以用公式表示成

$$x_1=\frac{-p+\sqrt{p^2-4q}}{2}, \qquad x_2=\frac{-p-\sqrt{p^2-4q}}{2},$$

我们假定 $p>0$, $q>0$, $4q\ll p^2$. 从表面来看，计算公式已经有了，我们只需要把公式中有关各项分别加以计算就可以了. 这种盲目的作法会对计算结果带来严重影响. 今分析于下：因为

$$x_2=\frac{-p-\sqrt{p^2-4q}}{2}=-\frac{p+\sqrt{p^2-4q}}{2}.$$

而 p 又为正数，所以这实际上在右端只要把 $\sqrt{p^2-4q}$ 计算出来，再执行一次加法运算就可以了. 分母中的 2，在二进制计算机上相当于移位，容易实现. 但是，由于 $4q\ll p^2$，而 q 又为正数，所以 $\sqrt{p^2-4q}\approx p$. 因此，在利用前述公式来实现 x_1 的计算就会产生严重的后果. 因为分子的计算是两个相近的数相减，其结果会产生很大的相对误差！这种情况在考虑算法时必须予以重视，设法避免. 对本例的具体情况，我们可以采用两种方法来克服上述困难. 第一种办法是：计算 x_1 的公式等价于

$$x_1=\frac{-4q}{2(-p-\sqrt{p^2-4q})}=\frac{2q}{p+\sqrt{p^2-4q}}.$$

由于在计算 x_2 时，$p+\sqrt{p^2-4q}$ 已经计算出来了，只需再执行一次除法运算就可以了，这就避免了相近数相减这种不利情况. 我们顺便指出，数学上的所谓等价表达式，在实际计算

中是并不等价的，不同表达式的计算结果可能会有很大差异．当然，这里的运算是指在仅具有限字长的计算机上进行的．第二种处理方法是，在 x_2 计算出以后，根据根与系数的关系 $x_1x_2=q$，从而有 $x_1=q/x_2$．作一次除法就把 x_1 计算出来了．

所谓算法，说得确切一些就是：一些数据，它们可以是事先给定的或者在计算过程中的中间结果，按照规定的顺序进行运算的一个序列．

建立一个好的算法并不是一件容易的事。在建立算法时，有许多事情应当予以注意，例如，原始数据和中间结果的存贮（对大型问题尤其如此），运算量的大小（特别是乘、除法的运算次数，不同的算法也会有很大差别），在什么时候应当使用多倍精度运算以及如何充分发挥计算机本身的功能．此外，对于计算数学工作者来说，一件十分重要的事是：如何避免在计算过程中舍入误差的积累！对同一个计算问题，不同的算法，由于舍入误差积累而产生的影响也会有很大差别．前述二次代数方程求根的例，就说明了这一点．这个问题就是算法的数值稳定性问题．即使一个非常良态的问题，如果使用算法不当，也可以使计算结果遭到完全破坏，使其变得毫无意义．为了说明这种危害，今再举一例．

线性方程组

$$\begin{cases} 0.0001x+y=1 \\ x+y=2 \end{cases}$$

是非常良态的，即系数的微小摄动对解只产生很小的影响．但是，如果我们使用 Gauss 消去法而按照自然顺序选主元在十进制三位浮点数的机器上进行计算，具体情况如下：

以第一方程中 x 的系数 0.0001 为主元从第二个方程中消去 x 得

$$-10000y = -10000, \text{ 即 } y = 1.$$

代入第一个方程得

$$0.0001x + 1 = 1,$$

即 $x = 0$. 而如果以第二方程中 y 的系数 1 为主元，从第一个方程中消去 y 便得到

$$x = 1.$$

代入第二个方程就得到 $y = 1$. 它们和准确解 $x = 1.0001$, $y = 0.9999$ 相比，准确到三位有效数字.

以上二例都说明对同一个计算问题，使用的算法不同，效果也大不一样，于是就产生了关于算法的**数值稳定性问题**.

算法的数值稳定性和问题的病态性一样，都是就互相对比而言的，从这种意义来讲它们都带有相对的性质. 如何来判断一个算法的数值稳定性的好坏，需要应用舍入误差分析的方法. 关于舍入误差分析，一般说来有两种途径. 一种是所谓**向前误差分析**，即根据计算机中浮点算术运算的舍入误差传播规律，在算法执行的过程中逐步作出估计. 这种方法，一方面十分费事，同时，逐次所作的估计也比较保守（当然也可以应用统计的方法加以改善）. 另一种方法，就是由 Wilkinson 所提出的**向后误差分析**方法. 对于一般的线性问题，应用比较方便. 现在，我们以线性方程组为例，来简略说明向后误差分析的基本思想，并在此基础上，说明算法的数值稳定性的意义和判断方法.

设给定求解的线性方程组为

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{a}$$

其中 A 是一个 $n \times n$ 阶非奇异矩阵，$\boldsymbol{b}$ 为给定的 n 维向量，$\boldsymbol{x}$ 为待定的 n 维向量. 应用某种算法，例如 Gauss 主元素消去法求解时，我们得到的解 $\tilde{\boldsymbol{x}}$，它仅是方程组 (a) 的一个近似解.

然而，我们却可以根据浮点运算的误差传播规律，对这个算法的过程加以分析，最后可以得到这样的结论，我们计算得到的近似解 $\tilde{\boldsymbol{x}}$，实际上是某一个摄动方程

$$(A+\delta A)\tilde{\boldsymbol{x}}=\boldsymbol{b}+\delta\boldsymbol{b} \tag{b}$$

的准确解．在这里起重要作用的是 δA 的大小．我们可以对 δA 作出估计，不同的算法所产生的摄动，其大小也各不相同．$\|\delta A\|$ 的大小，就表示算法的数值稳定性的好坏．$\|\delta A\|$ 愈小，所用算法的数值稳定性也就愈好．这种方法，对矩阵求逆以及线性最小二乘问题求解，都同样适用．这是一种有效的误差分析方法．

1-3 病态程度的度量

在前面我们对一个计算问题的病态性质这个概念，作了概略的定性说明．现在我们来考虑如何对问题的病态程度作出定量估计．为了便于说明，我们把所考虑的计算问题表示成方程

$$f(\boldsymbol{x},\ \boldsymbol{\beta})=0 \tag{8}$$

的求解问题．其中的 $\boldsymbol{\beta}$ 为问题中的参数向量，$\boldsymbol{x}$ 是待定向量，我们假定 $\boldsymbol{x}$ 就是方程(8)的理论解．若参数 $\boldsymbol{\beta}$ 有摄动 $\delta\boldsymbol{\beta}$，我们设 $\boldsymbol{x}+\delta\boldsymbol{x}$ 为摄动问题

$$f(\boldsymbol{x}+\delta\boldsymbol{x},\ \boldsymbol{\beta}+\delta\boldsymbol{\beta})=0 \tag{9}$$

的理论解．因为已经有了范数概念，这就对我们考虑这样的问题带来很大方便．我们可以用 $\|\delta\boldsymbol{\beta}\|/\|\boldsymbol{\beta}\|$ 表示参数的相对摄动量(或称之为相对误差)，解的相对误差就可以近似地表示成 $\|\delta\boldsymbol{x}\|/\|\boldsymbol{x}\|$ 或 $\|\delta\boldsymbol{x}\|/\|\boldsymbol{x}+\delta\boldsymbol{x}\|$．如果我们能够对解的相对摄动作出估计，例如

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}+\delta\boldsymbol{x}\|}\approx c\frac{\|\delta\boldsymbol{\beta}\|}{\|\boldsymbol{\beta}\|}，设\quad \|\boldsymbol{\beta}\|\neq 0, \tag{10}$$

右端的 c 是一个正常数，它表示参数的相对摄动量对解的相对摄动量影响的程度，或者说它是相对误差的放大率．常数因子 c 的大小就是问题病态程度的反映．因此，我们便称因子 c 是问题(8)的**条件数**．

应当指出，在这里我们所以要考虑相对摄动而不直接考虑绝对摄动 $\|\delta\boldsymbol{\beta}\|$ 和 $\|\delta\boldsymbol{x}\|$，其原因在于只有相对摄动才能反映计算结果的精度．

我们还要指出，求得近似等式(10)，一般是非常困难的，实际上，对于一般的线性问题，我们能够得到估计式：

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}+\delta\boldsymbol{x}\|}\leqslant c\frac{\|\delta\boldsymbol{\beta}\|}{\|\boldsymbol{\beta}\|}. \tag{11}$$

现在我们以线性方程组(a)的求解为例，来给出条件数的具体估计．

根据第3章 **4-1** 中的公式(d)，当 $\|A^{-1}\|\|\delta A\|<1$ 时

$$B^{-1}-A^{-1}=-B^{-1}(\delta A)A^{-1},\ B=A+\delta A,$$

并假定 $\boldsymbol{b}$ 无摄动，两端右乘以 $\boldsymbol{b}$ 就得到

$$\|\delta\boldsymbol{x}\|\leqslant\|B^{-1}\|\|\delta A\|\|\boldsymbol{x}\|,$$

或

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\frac{\|\delta A\|\|A^{-1}\|}{1-\|\delta A\|\|A^{-1}\|}$$

$$=\|A\|\|A^{-1}\|\frac{1}{1-\|\delta A\|\|A^{-1}\|}\frac{\|\delta A\|}{\|A\|}. \tag{12}$$

如果 $\|\delta\boldsymbol{x}\|$ 与 $\|\boldsymbol{x}\|$ 相比很小，则从

$$(A+\delta A)(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b}$$

立刻可以推得与(12)近似的不等式

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}+\delta\boldsymbol{x}\|}\leqslant\|A\|\|A^{-1}\|\frac{\|\delta A\|}{\|A\|}. \tag{13}$$

在特殊情况下，不等式(13)可以变为等式．例如，根据矩阵范数的定义，总存在向量 $\boldsymbol{w}$ 使

$$\|A^{-1}\boldsymbol{w}\|=\|A^{-1}\|\|\boldsymbol{w}\|,\ \boldsymbol{w}\neq 0.$$

任意给定小常数 $\beta\neq 0$, 令 $\delta\boldsymbol{x}=-\beta A^{-1}\boldsymbol{w}$, $\boldsymbol{x}+\delta\boldsymbol{x}=\boldsymbol{w}$, $\boldsymbol{b}=(A+\beta I)\boldsymbol{w}$, $\delta A=\beta I$, 总有

$$A\boldsymbol{x}=\boldsymbol{b},\ (A+\delta A)(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b},$$

$$\|\delta\boldsymbol{x}\|=|\beta|\|A^{-1}\boldsymbol{w}\|=\|\delta A\|\|A^{-1}\|\|\boldsymbol{x}+\delta\boldsymbol{x}\|,$$

从而有

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}+\delta\boldsymbol{x}\|}=\|A^{-1}\|\|A\|\frac{\|\delta A\|}{\|A\|}. \tag{14}$$

只要 $-\beta$ 不是矩阵 A 的特征值,矩阵

$$A+\delta A=A+\beta I$$

总为非奇异.

其次,如果矩阵 A 没有摄动,而只有右端向量 $\boldsymbol{b}$ 有摄动 $\delta\boldsymbol{b}$, 设 $A\boldsymbol{x}=\boldsymbol{b}$, $A(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b}+\delta\boldsymbol{b}$, 可证不等式

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\|A^{-1}\|\|A\|\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|} \tag{15}$$

成立,在这里我们假定 $\boldsymbol{x}\neq 0$, $\boldsymbol{b}\neq 0$.

在特殊情形,(15)也会变成等式. 因为,总存在向量 $\delta\boldsymbol{b}\neq 0$, $\boldsymbol{x}\neq 0$, 使

$$\|A^{-1}\delta\boldsymbol{b}\|=\|A^{-1}\|\|\delta\boldsymbol{b}\|,\ \|A\boldsymbol{x}\|=\|A\|\|\boldsymbol{x}\|.$$

于是,因为 $A\delta\boldsymbol{x}=\delta\boldsymbol{b}$, 所以 $\delta\boldsymbol{x}=A^{-1}\delta\boldsymbol{b}$, 从而有

$$\|\delta\boldsymbol{x}\|=\|A^{-1}\|\|\delta\boldsymbol{b}\|,$$

显然有等式

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}=\|A^{-1}\|\|A\|\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}. \tag{16}$$

从以上的讨论可以看出,在(12),(13)和(15)中都出现了一个共同的因子 $\|A^{-1}\|\|A\|$, 这个因子就相当于(11)中的因子 c, 它反映参数的相对摄动对解的相对摄动的影响,因此我们有

定义 1 今规定 $\|A\|\|A^{-1}\|$ 为线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 求解

问题的**条件数**.

关于矩阵 A 的求逆问题，若假定 $\|A^{-1}\|\|\delta A\|<1$，则根据第3章**4-1**中的公式(d)

$$B^{-1}-A^{-1}=-B^{-1}(\delta A)A^{-1},\ B=A+\delta A,$$

立得
$$\|B^{-1}-A^{-1}\|\leqslant\|B^{-1}\|\|\delta A\|\|A^{-1}\|,$$

即

$$\frac{\|B^{-1}-A^{-1}\|}{\|A^{-1}\|}\leqslant\frac{\|\delta A\|\|A^{-1}\|}{1-\|\delta A\|\|A^{-1}\|}$$

$$=\frac{\|A^{-1}\|\|A\|}{1-\|\delta A\|\|A^{-1}\|}\frac{\|\delta A\|}{\|A\|}. \tag{17}$$

由此可以看出，因子 $\|A^{-1}\|\|A\|$ 仍为相对摄动的放大因子，所以，$\|A^{-1}\|\|A\|$ 也是非奇异矩阵求逆的条件数. 或者简称之为**矩阵 A 的条件数**，记作

$$k(A)=\|A^{-1}\|\|A\|. \tag{18}$$

因为本书中一般只使用矩阵的谱范数，所以，上述条件数又可以称为矩阵 A 的**谱条件数**. 根据矩阵的范数和它的奇异值之间的关系，$k(A)$ 的表达式(18)可以表示成

$$k(A)=\frac{\lambda_1}{\lambda_n}, \tag{19}$$

其中 λ_1 和 λ_n 分别为非奇异矩阵的最大和最小奇异值.

第一小节方程(4)的系数矩阵 A 是正定的，它的最大和最小特征值分别为：$\lambda_1\approx30.2887$，$\lambda_4\approx0.01015$，可知 $k(A)=\frac{\lambda_1}{\lambda_4}\approx2984=2.984\times10^3$.

1-4 病态程度的度量(续)

在前一小节我们已经对一般的线性方程组求解以及非奇异矩阵求逆问题的病态性程度作出了估计，并建立相应于这两个问题的条件数概念及其表达式. 在这一小节中我们将就

广义逆矩阵 A^+ 的计算问题以及与其相关联的最小二乘问题的病态性程度给出相应的条件数．根据前一章中 **4-1** 中的公式(37)，对一般的 $m\times n$ 阶矩阵 A 和 B，有估计式

$$\|B^+-A^+\|\leqslant(\|A^+\|\|B^+\|+\|B^+\|^2+\|A^+\|^2)\|B-A\|.$$

现在我们令 $B=A+\delta A$，并假定 $\mathrm{rank}(B)=\mathrm{rank}(A)$，并且 $\|A^+\|\|\delta A\|<1$，则有

$$\|B^+-A^+\|\leqslant\|A^+\|\left(\frac{\|A^+\|}{1-\Delta}+\frac{\|A^+\|}{(1-\Delta)^2}+\|A^+\|\right)\|\delta A\|.$$

从而便得到相对摄动的估计式：

$$\frac{\|B^+-A^+\|}{\|A^+\|}\leqslant(1+\eta+\eta^2)\|A^+\|\|A\|\frac{\|\delta A\|}{\|A\|},\tag{20}$$

其中

$$\eta=\frac{1}{1-\Delta},\quad \Delta=\|A^+\|\|\delta A\|.\tag{21}$$

如果 $\|\delta A\|$ 能使 $\|A^+\|\|\delta A\|\leqslant\frac{1}{2}$，则不等式(20)就化为

$$\frac{\|B^+-A^+\|}{\|A^+\|}\leqslant 7\|A^+\|\|A\|\frac{\|\delta A\|}{\|A\|}.\tag{22}$$

不论从估计式(20)或(22)都可看出，因子 $\|A^+\|\|A\|$ 都是反映计算广义逆矩阵 A^+ 时病态程度的度量．仿前一小节，我们有

定义 2 我们规定 $\|A^+\|\|A\|$ 为计算广义逆矩阵 A^+ 问题的条件数，并仍记之为

$$k(A)=\|A^+\|\|A\|.\tag{23}$$

根据矩阵范数和奇异值之间的关系，设 $\mathrm{rank}(A)=k$，则条件数(23)可以用 A 的最大奇异值 λ_1 和最小奇异值 λ_k 表示成：

$$k(A)=\frac{\lambda_1}{\lambda_k}.\tag{24}$$

所以，对于计算广义逆矩阵 A^+ 的条件数，也称之为矩阵 A 的

谱条件数。显然，如果 A 为方阵并且非奇异，(24)和(19)相同。

现在我们来考虑相应的线性最小二乘问题，即求超定线性方程 $A\boldsymbol{x}=\boldsymbol{b}$ 的极小最小二乘解的问题。今假定 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 都有摄动，摄动问题的解为 $\boldsymbol{x}+\delta\boldsymbol{x}$，即

$$\boldsymbol{x}+\delta\boldsymbol{x}=(A+\delta A)^{+}(\boldsymbol{b}+\delta\boldsymbol{b}).$$

根据前章§6公式(59)，设

$$\operatorname{rank}(A+\delta A)=\operatorname{rank}(A),\quad \|A^{+}\|\|\delta A\|<1,$$

我们有

$$\|\delta\boldsymbol{x}\|\leqslant\frac{\|A^{+}\|}{1-\Delta}\left(2\|\delta A\|\|\boldsymbol{x}\|+\|\delta\boldsymbol{b}\|+\frac{\|\delta A\|\|A^{+}\|\|\boldsymbol{r}\|}{1-\Delta}\right),$$

据此可得

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\frac{k(A)}{1-\Delta}\left[\left(2+\frac{k(A)\|\boldsymbol{r}\|}{\|A\boldsymbol{x}\|}\right)\frac{\|\delta A\|}{\|A\|}+\frac{\|\boldsymbol{b}\|}{\|A\boldsymbol{x}\|}\frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}\right],\tag{25}$$

其中，$r=\boldsymbol{b}-A\boldsymbol{x}$.

和前一小节的情形完全一样，$k(A)=\|A\|\|A^{+}\|$ 也反映最小二乘问题求解的条件。

关于矩阵条件数的若干简单性质

非奇异方阵求逆和相应的线性方程组求解问题是一般的 $m\times n$ 阶矩阵 A 计算广义逆矩阵 A^{+} 和相应的最小二乘问题求解问题的特例，它们的条件数也同样如此。现在我们就后者列举关于条件数 $k(A)$ 的若干简单性质如下，这里我们假定 A 为 $m\times n$ 阶，$m\geqslant n$，$\operatorname{rank}(A)=k$，A的最大和最小奇异值分别为 λ_1 和 λ_k.

1° $k(A)\geqslant1$，因为 $\lambda_1\geqslant\lambda_k$，$k(A)=\lambda_1/\lambda_k$.

2° 对任何 $\alpha\neq0$，$k(\alpha A)=k(A)$. 因为 αA 的最大和最

小奇异值分别为 $|\alpha|\lambda_1$ 和 $|\alpha|\lambda_k$, 所以 $k(\alpha A)=|\alpha|\lambda_1/(|\alpha|\lambda_k)$ $=\lambda_1/\lambda_k=k(A)$.

3° 在直交变换下矩阵 A 的条件数不变，即对任何 $m\times m$ 阶直交阵，恒有 $k(QA)=k(A)$.

因为 $(QA)^T(QA)=A^TA$, 所以 QA 的奇异值和 A 的完全相同.

4° 设 Q_k 为任意的 $m\times k$ 阶列直交矩阵，$m\geq k$, 则 $k(Q_k)$ $=1$.

因为 $Q_k^TQ_k=I_k$, 所以 Q_k 的 k 个奇异值都是 1.

1-5 奇异性和病态概念的区别

在本节的末了，我们打算对奇异性和病态这两个概念作进一步的阐述. 首先我们要指出，奇异性和病态性这两个概念分属于两种完全不同的范畴. 例如，对方阵 A 来说，奇异性是和通常意义下的逆矩阵 A^{-1} 是否存在相联系的. 如果逆矩阵 A^{-1} 不存在，便称这样的矩阵是奇异的. 说得更具体一些，就是如果存在同阶方阵 B, 它具有性质: $BA=AB=I$, 我们就称这样的矩阵 A 是非奇异的，B 就称之为 A 的逆矩阵. 同时，奇异性又是某种意义下的间断性. 这种间断性表现为:

设 $\{A_i\}$ 是一个非奇异矩阵序列，其极限为奇异矩阵 A. 因为 A_i 非奇异，所以对 $i=1, 2, \cdots$, 逆矩阵 A_i^{-1} 都存在，但是极限

$$\lim_{i\to\infty} A_i^{-1}$$

却不存在. 如果把 A_i 当作一个抽象的变量，把 A_i^{-1} 当作这个变量的函数 $f(A_i)$, 那么，当 $A_i\to A$ 时，函数 $f(A_i)$ 的极限不存在，而且这个函数在 A 点根本就没有定义. 不但如此，无论如何规定 $f(A_i)$ 在 A 点的值，都无法使它在 A 点连续. 换句话说，A 还是 $f(A_i)$ 的一种不可移间断点.

至于病态性这个概念，这是针对某一个计算问题而言的，它是这个计算问题的一种属性．就矩阵 A 的求逆问题（在通常的意义下）来说，如果是奇异的，从逻辑上来说，就根本不会有求逆问题．人们决不会去寻求明知是理论上不存在的东西，只有矩阵 A 为非奇异，才会发生它的求逆问题．计算 A^{-1} 这个问题的病态程度如何，说得通俗一些，就表现在关于 A^{-1} 的计算是否容易实现．

其次，关于以 A 为系数矩阵的线性方程组（在通常意义下）的求解问题，在前面我们考虑这种计算问题的病态性质的时候，我们曾假设矩阵为非奇异的，而摄动矩阵 $A+\delta A$ 也是非奇异的，从表面看来，似乎方程组求解问题的病态性质和矩阵 A 的非奇性有所联系．这仅是一种表面现象。我们之所以要作出关于矩阵 A 的非奇异性的假定，是由于这一类计算问题是大量存在的．但是，即使系数矩阵 A 为奇异，仍然存在着这样的线性方程组的求解问题，甚至是在通常意义下的求解问题．当然，我们要假定这个方程组是相容的．虽然系数矩阵 A 为奇异，但它和这个求解问题是否病态全无关系．相应的求解问题可能是病态的，也可能是良态的，究竟如何，可以从 $\|A^+\|\|A\|$ 的大小来判断，而 A^+ 的存在是和矩阵 A 是否奇异没有关系的．

§2　数值相关性理论

2-1　引言

在前一节中，我们已经对数学计算问题的病态性质这个具重要实际意义的概念作了简要的阐述．我们的讨论虽然是针对本书所涉及的有关有限维线性问题进行的，但是它的实

际意义和作用却更为广泛．一方面，计算问题的病态理论，对线性问题所得到的结果，对于一般的非线性问题，在研究算法的理论时，也同样重要．众所周知，处理非线性问题的一个有普遍意义的原则，就是**逐次线性化**．例如牛顿迭代法是解非线性方程组的一类重要方法，在迭代过程中的每一步，都遇到一个以 Jacobi 矩阵为系数矩阵的线性方程组的求解问题，而相应的 Jacobi 矩阵的病态程度，对牛顿法的实际效果，有不可忽视的影响．如何改善这种状态，仍是值得进一步研究的课题．在无约束非线性最优化方法中，拟牛顿法占有重要的地位．虽然这一类方法可以避免矩阵的直接求逆，而可以采取直接逼近 Hesse 矩阵逆的方法，但是，如果 Hesse 阵本身病态性严重的话，逼近它是非常困难的．另一方面，解大型稀疏线性问题是一类十分重要的课题，而迭代法对于解决存贮问题，具有明显的优点，关键问题是迭代法的收敛速度，而迭代法的收敛速度则决定于迭代矩阵的病态程度．要想提高收敛速度，就得设法改善迭代矩阵的条件．在五十年代发展起来的一类所谓超松弛法，实际上就是实现这种思想的一种具体方法．近年来，对条件予优(Pre-conditioning)的理论和方法的研究，日益受到人们的重视，并且已经收到明显的效果．我们所以要讲以上这一段话，其目的就在于说明对处理病态问题的理论和方法的研究，在计算数学的各个领域中都具有十分重要的实际意义．对于开展"处理病态问题的理论和方法"的研究，应当注意两个方面：其一是如何建立对病态性程度判断的有效方法，其二是在病态严重时如何加以处理以改善算法的效果．在这一章中我们着重讨论病态性程度的判断问题．

2·2 Gram 行列式的一些重要性质

相关性理论无论在纯粹数学的研究中以及在数值计算的

研究中都非常重要．但是，古典的线性相关性理论，有很大的局限性．为了适应数值计算研究的需要，必须进一步加以发展．在本章中我们将用定量的方法对线性相关性的古典理论进行改造，为研究病态问题的理论与应用提供必要基础．为了这一目的，我们首先对有关 Gram 行列式的一些重要性质作简要的介绍．

设 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 为 n 个 m 维列向量，并假定 $m>n$．记 A_k 为以 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为列组成的 $m\times k$ 阶矩阵，$k\leqslant n$，

$$A_k=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] \tag{26}$$

向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为线性相关的数学定义是：存在不全为零的常数 $c_1, \cdots, c_k$，使

$$c_1\boldsymbol{a}_1+c_2\boldsymbol{a}_2+\cdots+c_k\boldsymbol{a}_k=\mathbf{0}.$$

这就是说，一个向量系从理论上讲，不是线性相关就是线性独立．从线性独立到线性相关没有一个量变过程．它是用一刀切的方式来定义相关性这个概念的．从纯数学的角度来看，这样的定义是非常清楚的．但是，从数值计算的角度来看，这个概念往往就会变得模糊不清．从这个定义出发，用数值计算的手段要想判断一个给定的向量系是否为线性相关，常常难以实现．这就表现出古典的线性相关概念，存在着严重的缺陷．本节将为以后采用定量的方法对古典的线性相关性这一重要概念进行改造提供必要的基础．

根据古典的线性相关性理论，向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为线性相关的充分必要条件是这个向量系的 Gram 行列式为零，

$$\det(A_k^T A_k)=0, \tag{27}$$

其中，$k\times k$ 阶矩阵 $A_k^T A_k$ 的 (i, j) 元素为

$$a_{ij}=\boldsymbol{a}_i^T\boldsymbol{a}_j,\ i, j=1, 2, \cdots, k.$$

为了用定量的方法来研究线性相关性的数值理论，我们

有必要对向量系的Gram行列式的若干重要性质加以分析.

为便于讨论，我们把向量系 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 的 Gram 行列式用记号表示成

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]. \tag{28}$$

首先我们假定向量系 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 是线性独立的，从而就保证了它的任何子系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 也是线性独立的. 记 S_k 为向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 所张成的 k 维子空间,

$$S_k = \mathrm{Span}\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k\}. \tag{29}$$

然后作空间 $\mathscr{R}^m$ 的直交分解

$$\mathscr{R}^m = S_k \oplus S_k^{\perp}. \tag{30}$$

于是，空间 $\mathscr{R}^m$ 中的任何向量 $\boldsymbol{x}$ 都可以唯一地表示成:

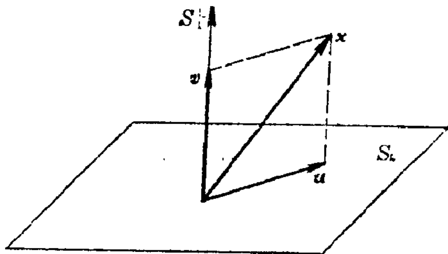


图 4-1

$$\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v},\quad \boldsymbol{u} \in S_k,\quad \boldsymbol{v} \in S_k^{\perp}. \tag{31}$$

向量 $\boldsymbol{v}$ 就是向量 $\boldsymbol{x}$ 在 $S_k^{\perp}$ 上的直交投影，其长度 $\|\boldsymbol{v}\|$ 就是向量 $\boldsymbol{x}$ 到子空间 S_k 的最短距离. 记 $h = \|\boldsymbol{v}\|$. 今确定 h 的值于下.

显然，向量 $\boldsymbol{v}$ 是 $\boldsymbol{x}$ 在 $S_k^{\perp}$ 上的直交投影，所以可用直交投影矩阵把 $\boldsymbol{v}$ 表示成:

$$\boldsymbol{v} = (I - A_k A_k^+)\boldsymbol{x}.$$

于是

$$\|\boldsymbol{v}\|^2 = \boldsymbol{x}^T (I - A_k A_k^+)\boldsymbol{x}. \tag{32}$$

因为

$$\mathrm{rank}(A_k) = k,$$

所以

$$A_k^+ = (A_k^T A_k)^{-1} A_k^T.$$

令

$$\boldsymbol{y} = (A_k^T A_k)^{-1} A_k^T \boldsymbol{x}, \tag{33}$$

则利用 Cramer 法则，可以把向量 $\boldsymbol{y}$ 的分量 $\eta_1, \cdots, \eta_k$ 表示成：

$$\eta_j = \frac{1}{G_k} \cdot \det\begin{bmatrix} (\boldsymbol{a}_1, \boldsymbol{a}_1)\cdots(\boldsymbol{a}_1, \boldsymbol{a}_{j-1})(\boldsymbol{a}_1, \boldsymbol{x})(\boldsymbol{a}_1, \boldsymbol{a}_{j+1})\cdots(\boldsymbol{a}_1, \boldsymbol{a}_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ (\boldsymbol{a}_k, \boldsymbol{a}_1)\cdots(\boldsymbol{a}_k, \boldsymbol{a}_{j-1})(\boldsymbol{a}_k, \boldsymbol{x})(\boldsymbol{a}_k, \boldsymbol{a}_{j+1})\cdots(\boldsymbol{a}_k, \boldsymbol{a}_k) \end{bmatrix}$$

$$= \frac{(-1)^{k-j}}{G_k} \cdot \det\begin{bmatrix} (\boldsymbol{a}_1, \boldsymbol{a}_1)\cdots(\boldsymbol{a}_1, \boldsymbol{a}_{j-1})(\boldsymbol{a}_1, \boldsymbol{a}_{j+1})\cdots(\boldsymbol{a}_1, \boldsymbol{a}_k)(\boldsymbol{a}_1, \boldsymbol{x}) \\ (\boldsymbol{a}_k, \boldsymbol{a}_1)\cdots(\boldsymbol{a}_k, \boldsymbol{a}_{j-1})(\boldsymbol{a}_k, \boldsymbol{a}_{j+1})\cdots(\boldsymbol{a}_k, \boldsymbol{a}_k)(\boldsymbol{a}_k, \boldsymbol{x}) \end{bmatrix}$$

$$= (-1)^{k-j}\Delta_j / G_k \quad (\text{设行列式的值为 } \Delta_j)$$

其中 $G_k = G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]$. 从而便有

$$\|\boldsymbol{v}\|^2 = \boldsymbol{x}^T\boldsymbol{x} - \boldsymbol{x}^T A_k A_k^+ \boldsymbol{x}_k$$

$$= \frac{1}{G_k}\left[(\boldsymbol{x}, \boldsymbol{x})G_k + \sum_{j=1}^{k} (-1)^{k-j+1}(\boldsymbol{a}_j, \boldsymbol{x})\Delta_j\right]$$

$$= \frac{1}{G_k}\det\begin{bmatrix} & (\boldsymbol{a}_1, \boldsymbol{x}) \\ A_k^T A_k & \vdots \\ & (\boldsymbol{a}_k, \boldsymbol{x}) \\ (\boldsymbol{a}_1, \boldsymbol{x})\cdots(\boldsymbol{a}_k, \boldsymbol{x}) & (\boldsymbol{x}, \boldsymbol{x}) \end{bmatrix}$$

$$= G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}] / G_k, \tag{34}$$

即

$$h^2 = G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}] / G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k],$$

或

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}] = h^2 G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]. \tag{35}$$

据此便得到下列关系：

$$\left.\begin{aligned} &G[\boldsymbol{a}_1, \boldsymbol{a}_2] = \|\boldsymbol{v}_1\|^2 G[\boldsymbol{a}_1] = \|\boldsymbol{v}_1\|^2 \|\boldsymbol{a}_1\|^2, \\ &G[\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3] = \|\boldsymbol{v}_2\|^2 G[\boldsymbol{a}_1, \boldsymbol{a}_2], \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] = \|\boldsymbol{v}_{k-1}\|^2 G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}], \end{aligned}\right\} \tag{36}$$

其中的向量 $\boldsymbol{v}_j(j=1, \cdots, k-1)$ 表示 $\boldsymbol{a}_{j+1}$ 在 $S_j^\perp$ 上的直交投影。

从 (36) 可以看出，Gram 行列式具有明显的几何意义：

$G[\boldsymbol{a}_1]=\|\boldsymbol{a}_1\|^2$，这就表示 $G[\boldsymbol{a}_1]$ 是向量 $\boldsymbol{a}_1$ 的长度平方。$G[\boldsymbol{a}_1, \boldsymbol{a}_2]=\|\boldsymbol{v}_1\|^2\|\boldsymbol{a}_1\|^2$，表示以 $\boldsymbol{a}_1, \boldsymbol{a}_2$ 为两边的平行四边形的面积的平方，而 $G[\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]=\|\boldsymbol{v}_2\|^2G[\boldsymbol{a}_1, \boldsymbol{a}_2]$ 就表示以向量 $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ 为三条棱的平行六面体的体积的平方，余类推。

根据上述讨论，我们便不难推出 Gram 行列式的下列重要性质：我们总假定 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 为线性独立，于是有

1° 对于 $k=1, 2, \cdots, n$，恒有

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]>0.$$

2° 设 $p_1, p_2, \cdots, p_k$ 为自然数列 $1, 2, 3, \cdots, k$ 的任何一种排列，则恒有

$$G[\boldsymbol{a}_{p_1}, \boldsymbol{a}_{p_2}, \cdots, \boldsymbol{a}_{p_k}]=G[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k],$$

这就是说，任意交换诸 $\boldsymbol{a}_i$ 的次序，Gram 行列式的值不变。

因为这种向量次序的交换，相当于在行列式

$$\det(A_k^TA_k)$$

中的相应的行与行之间和列与列之间的交换，由于交换次数恒为偶数，所以行列式的值不变。

3° $G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant G[\boldsymbol{a}_1]G[\boldsymbol{a}_2]\cdots G[\boldsymbol{a}_k]$.

这个不等式可以根据 (36) 直接推导出来。因为 $\boldsymbol{v}_1$ 是向量 $\boldsymbol{a}_2$ 在 $\boldsymbol{a}_1$ 所张 1 维子空间的直交补空间上的直交投影，所以 $\|\boldsymbol{v}_1\|\leqslant\|\boldsymbol{a}_2\|$. 从而有 $G[\boldsymbol{a}_1, \boldsymbol{a}_2]\leqslant\|\boldsymbol{a}_2\|^2G[\boldsymbol{a}_1]=G[\boldsymbol{a}_1]G[\boldsymbol{a}_2]$. 利用归纳法便不难证明不等式 3°.

4° 上述不等式可以推广到更为一般的形式：

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r]G[\boldsymbol{a}_{r+1}, \cdots, \boldsymbol{a}_k],$$

其中 $1\leqslant r\leqslant k$.

5° 若 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 互相直交，则 3° 变成等式：

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]=G[\boldsymbol{a}_1]G[\boldsymbol{a}_2]\cdots G[\boldsymbol{a}_k].$$

若向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r$ 和向量系 $\boldsymbol{a}_{r+1}, \cdots, \boldsymbol{a}_k$ 相互直交，即

$$(\boldsymbol{a}_i, \boldsymbol{a}_j)=0, \text{对 } i=1, 2, \cdots, r; j=r+1, \cdots, k,$$

则 4° 变为等式：

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]=G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r]G[\boldsymbol{a}_{r+1}, \cdots, \boldsymbol{a}_k].$$

6° 设 $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k$ 为一直交系，并且

$$\|\boldsymbol{b}_i\|=\|\boldsymbol{a}_i\|, i=1, 2, \cdots, k,$$

则恒有
$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant G[\boldsymbol{b}_1, \cdots, \boldsymbol{b}_k],$$

并且等号只有在 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 也是一个直交系时才成立.

这就是说，在长度相同的向量系中，以直交系的 Gram 行列式的值为最大.

道理很简单，因为 $G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant G[\boldsymbol{a}_1]\cdots G[\boldsymbol{a}_k]$，如果 $\{\boldsymbol{a}_i\}$ 为直交系，则等式成立．反过来，若有任何两个向量不直交，据(36)就不难证明等式不能成立.

2-3 总体相关性和相对相关性

现在，我们便来着手在相关性定性理论的基础上进行定量化的工作．为此目的，我们首先对向量系的线性相关性程度建立一种明确的定量表示方法．今假定我们所考虑的向量系为 $\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n\}$，$\boldsymbol{a}_i$ 都是 m 维向量，且 $m\geqslant n$．第一步工作，是对这个向量系进行规范化，即对向量系中的向量都乘以一个共同的因子，使

$$\max_{1\leqslant k\leqslant n}\|\boldsymbol{a}_k\|=1.$$

这时，显然有

$$0\leqslant G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant 1, \ k=1, 2, \cdots, n, \tag{37}$$

并且，只有在向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为一标准直交系时，Gram 行

列式 $G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]$ 才能等于 1，在这个向量系为线性相关时，Gram 行列式的值才能为零.

(I) 总体相关性

定义 3 设 ε 为给定的一个正小数，$0<\varepsilon<1$，若

$$G^{\frac{1}{2}}[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] \leqslant \varepsilon,$$

我们便称向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为 **ε-线性相关**. 若对于某一正数 δ，$0<\delta\leqslant 1$，有

$$G^{\frac{1}{2}}[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] \geqslant \delta,$$

便称向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为 **δ-线性独立**.

根据这个定义，显然可知：

1) 若向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为 ε-线性相关，则它的任何一个扩张系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_r, r>k$，也必为 ε-线性相关.

2) 若向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为 δ-线性独立，则它的任何子系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_p$, $p<k$，也必为 δ-线性独立.

3) 向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 为 ε-线性相关或 δ-线性独立和向量的排列顺序无关.

定义中的正数 ε 和 δ 就是向量系(就其总体而言) $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 的线性相关性程度或线性独立性程度的一种度量.

定理 1 在直交变换下，向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k$ 的相关性程度(或独立性程度)不变.

证明 设 Q 为任意的 $m\times m$ 阶直交阵，则

$$[Q\boldsymbol{a}_1, \cdots, Q\boldsymbol{a}_k] = Q[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] = QA_k,$$

从而有 $$(QA_k)^T(QA_k) = A_k^T Q^T Q A_k = A_k^T A_k.$$

这就证明了

$$G[QA_k] = G[A_k]. \text{ 证毕.}$$

附注 为了简单起见，我们把 $G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]$ 记成 $G[A_k]$，把 $G[Q\boldsymbol{a}_1, \cdots, Q\boldsymbol{a}_k]$ 记成 $G[QA_k]$.

定理 2 设 $\beta_1, \beta_2, \cdots, \beta_k$ 为给定的正数，在满足关系

$$\|\boldsymbol{a}_i\|=\beta_i,\ i=1,\ \cdots,\ k$$

的一切向量系 $\{(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k)\}$ 中，直交系具有最强的线性独立性.

因为，在定理的条件下，只有直交系才使 Gram 行列式的值为最大.

定理 3 在向量系规范化的情况下，

$$G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k] \leqslant \min_{1\leqslant i\leqslant k}\|\boldsymbol{a}_i\|,\ k=1,\ 2,\ \cdots,\ n. \tag{38}$$

证明很简单，从略.

(II) 相对相关性

首先我们假定向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$ 为线性独立. 对于给定的向量 $\boldsymbol{x}$，现在我们来考虑它关于向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k (1\leqslant k\leqslant n)$ 的线性相关性程度的度量问题. 为了这个目的，我们来考虑用 k 维子空间

$$S_k=\mathrm{Span}\{\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k\}$$

中的向量逼近向量 $\boldsymbol{x}$ 的问题：确定系数 $c_1, c_2, \cdots c_k$ 使

$$\left\|\boldsymbol{x}-\sum_{i=1}^{k} c_i\boldsymbol{a}_i\right\|=\text{极小}. \tag{39}$$

根据我们在前面关于 Gram 行列式的几何性质的讨论可知，若令

$$\boldsymbol{v}=\boldsymbol{x}-\sum_{i=1}^{k} c_i\boldsymbol{a}_i, \tag{40}$$

欲使 $\|\boldsymbol{v}\|$ 为极小，则 $\boldsymbol{v}$ 应为向量 $\boldsymbol{x}$ 在 S_k 的直交补空间 $S_k^{\perp}$ 上的直交投影. 从而，相应的向量

$$\boldsymbol{u}=\sum_{i=1}^{k} c_i\boldsymbol{a}_i \tag{41}$$

便是 $\boldsymbol{x}$ 在子空间 S_k 上的直交投影. 这时 $\boldsymbol{v}$ 的长度平方为

$$\|\boldsymbol{v}\|^2=G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}]/G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]. \tag{42}$$

定义 4 我们规定 $\|\boldsymbol{v}\|$ 的极小值

$$\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}]/G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\}^{\frac{1}{2}}$$

为向量 $\boldsymbol{x}$ 关于向量系 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_k$ 的相对相关性指标，并记之为

$$E[\boldsymbol{x}|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]=\left\{\frac{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}]}{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]}\right\}^{\frac{1}{2}}. \tag{43}$$

附注 在考虑向量系的线性相关性度量时，应先把有关向量系规范化，这里的向量 $\boldsymbol{x}$，其长度应假定不大于1，即，$|\boldsymbol{x}|\leqslant 1$.

显然可知，若 $E[\boldsymbol{x}|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]=0$，则向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{x}$，在古典的意义下为线性相关．并且，根据 Gram 行列式的性质可知

$$0\leqslant E\{\boldsymbol{x}|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant\|\boldsymbol{x}\|\leqslant 1. \tag{44}$$

定理 4 若对某一个 i, $k<i\leqslant n$, 有

$$E[\boldsymbol{a}_i|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant\varepsilon,$$

则恒有

$$\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]\}^{\frac{1}{2}}\leqslant\varepsilon. \tag{45}$$

这就是说，若向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$ 中有一个向量 $\boldsymbol{a}_i$ 关于它的一个子系为 ε-线性相关，则该向量系就总体来说，也必为 ε-线性相关.

证明 因为

$$\begin{aligned}
&\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]\}^{\frac{1}{2}}\\
&\leqslant\frac{\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k, \boldsymbol{a}_i]G[\boldsymbol{a}_{k+1}, \cdots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_{i+1}, \cdots, \boldsymbol{a}_n]\}^{\frac{1}{2}}}{\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\}^{1/2}}\\
&\quad\times\{G[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\}^{\frac{1}{2}}\\
&\leqslant E[\boldsymbol{a}_i|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant\varepsilon.
\end{aligned}$$

从而定理得证．证毕.

§3 计算相关性指标的方法

3-1 引言

在进一步讨论线性相关性的定量理论在病态问题研究中的应用之前，我们先来介绍几种计算相关性指标的实用方法。从表面上看，要计算向量系的相关性指标，得计算 Gram 行列式的值，而实际上则并不需要单独进行相关性指标的计算。对于一些计算线性问题的重要算法，常常在算法的执行过程中，自然地计算出来，从而对算法的继续进行，提供有益的信息。

3-2 QR 分解法

对于给定的 $m\times n$ 阶矩阵 A，总存在一个 $m\times m$ 阶的直交阵 Q，使

$$Q^TA=R,$$

即

$$A=QR, \tag{46}$$

这里的 R 是一个 $m\times n$ 阶的上梯形阵，记其主对角元为

$$\rho_{11},\ \rho_{22},\ \cdots,\ \rho_{nn}. \tag{47}$$

我们称(46)为矩阵 A 的 QR 分解．它的具体实现，我们将在下一章中介绍．对矩阵 A 作适当的列交换，可以使

$$|\rho_{11}|\geqslant|\rho_{22}|\geqslant\cdots\geqslant|\rho_{nn}|. \tag{48}$$

若 $\mathrm{rank}(A)=n$，则 $\rho_{nn}\neq 0$．矩阵 R 的形式为

$$R=\begin{bmatrix}\widetilde{R}_n\\ 0\end{bmatrix}, \tag{49}$$

右端上方的 $\widetilde{R}_n$ 是一个 $n\times n$ 阶上三角阵，0 是 $(m-n)\times n$ 阶的零矩阵．如果把 $\widetilde{R}_n$ 左上角的那个 $k\times k$ 阶上三角阵记成 $\widetilde{R}_k$，那么，由于

$$A_k^TA_k=\widetilde{R}_k^T\widetilde{R}_k,\ k=1,\ 2,\ \cdots,\ n,\ A_n=A, \tag{50}$$

从而便得到关系

$$\det(A_k^T A_k)=\det(\tilde{R}_k^T)\det(\tilde{R}_k)$$
$$=\rho_{11}^2\rho_{22}^2\cdots\rho_{kk}^2,\ k=1,\ 2,\ \cdots,\ n. \tag{51}$$

据此便可以得到向量 $\boldsymbol{a}_k$ 关于向量系 $\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_{k-1}$ 的相对相关性指标：

$$E[\boldsymbol{a}_k|\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}]=|\rho_{kk}|. \tag{52}$$

应当指出，我们在定义向量系的相关性指标时，是假定向量系 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_n$ 已经规范化. 否则的话，上述公式应修正为

$$E[\boldsymbol{a}_k|\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}]=|\rho_{kk}|/|\rho_{11}|. \tag{53}$$

此外，还需要指出，为使序列 $\{|\rho_{ii}|\}$ 不增，$\boldsymbol{a}_1$ 应为向量系 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_n$ 中长度最大的向量，对规范化的向量系，$\|\boldsymbol{a}_1\|=1$，并且，因为在直交变换下向量的长度不变，这时应有 $|\rho_{11}|=1$.

3-3 Gram-Schmidt 直交化方法

根据向量 $\boldsymbol{a}_k$ 关于向量系 $\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \cdots,\ \boldsymbol{a}_{k-1}$ 相对相关性指标的定义，$E[\boldsymbol{a}_k|\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}]$ 实际上就是向量 $\boldsymbol{a}_k$ 在 Span $\{\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}\}$ 的直交补空间上的投影向量 $\boldsymbol{v}_k$ 的长度，而向量 $\boldsymbol{v}_k$ 可以表示成

$$\boldsymbol{v}_k=\boldsymbol{a}_k-\sum_{i=1}^{k-1}c_i\boldsymbol{a}_i,\ k=2,\ \cdots,\ n,\ \boldsymbol{v}_1=\boldsymbol{a}_1.$$

右端系数 $c_1,\ c_2,\ \cdots,\ c_{k-1}$ 应使 $\boldsymbol{v}_k$ 和 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}$ 都直交. 于是，向量系 $\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots,\ \boldsymbol{v}_n$ 必是一个直交系，并且，$\|\boldsymbol{v}_k\|$ 就是向量 $\boldsymbol{a}_k$ 关于向量系 $\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}$ 的相对相关性指标. Gram-Schmidt 直交化方法就是确定直交系 $\boldsymbol{v}_1,\ \cdots,\ \boldsymbol{v}_n$ 的一种方法. 其步骤如下：令 $\boldsymbol{v}_1=\boldsymbol{a}_1$，对 $k=2,\ 3,\ \cdots,\ n$，$\boldsymbol{v}_k$ 的表达式可以改写成

$$\boldsymbol{v}_k=\boldsymbol{a}_k-\sum_{i=1}^{k-1}\beta_i\boldsymbol{v}_i. \tag{54}$$

在 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}$ 依次确定以后，为确定 $\boldsymbol{v}_k$，可以选取系数 $\beta_1, \beta_2, \cdots, \beta_{k-1}$，使 $\boldsymbol{v}_k$ 和 $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_{k-1}$ 都直交就行了，这时，$\boldsymbol{v}_k$ 必然和 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_{k-1}$ 也都直交．而确定系数 β_i 要比确定前面的 $\boldsymbol{v}_k$ 的表达式中的系数 c_i 简便得多．因为 $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}$ 是一个直交系，所以

$$\beta_i=(\boldsymbol{a}_k, \boldsymbol{v}_i)/(\boldsymbol{v}_i, \boldsymbol{v}_i), \ i=1, 2, \cdots, k-1. \tag{55}$$

不难证明，$(\boldsymbol{v}_i, \boldsymbol{v}_i)=(\boldsymbol{v}_i, \boldsymbol{a}_i)$，所以 β_i 的表达式也可以写成

$$\beta_i=(\boldsymbol{a}_k, \boldsymbol{v}_i)/(\boldsymbol{v}_i, \boldsymbol{a}_i), \ i=1, 2, \cdots, k-1. \tag{56}$$

这就是通常所讲的古典的 Gram-Schmidt 直交化方法．

在六十年代，发现了这种直交化方法的数值稳定性不好，而它的一种变形却具有较好的数值稳定性．我们称这种方法为**改进的 Gram-Schmidt 直交化方法**．此法实现的步骤如下．

第一步：对 $i=2, 3, \cdots, n$，在向量 $\boldsymbol{a}_1$ 和 $\boldsymbol{a}_i$ 确定的平面上，确定向量 $\boldsymbol{a}_i^{(1)}$，使它们和 $\boldsymbol{a}_1$ 都直交，即

$$(\boldsymbol{a}_i^{(1)}, \boldsymbol{a}_1)=0, \ i=2, 3, \cdots, n.$$

并令 $\boldsymbol{a}_1^{(1)}=\boldsymbol{a}_1$．为具体确定 $\boldsymbol{a}_i^{(1)}$，令

$$\boldsymbol{a}_i^{(1)}=\boldsymbol{a}_1-r_i\boldsymbol{a}_i. \tag{57}$$

于是，据 $(\boldsymbol{a}_i^{(1)}, \boldsymbol{a}_1)=0$，便得到

$$r_i=(\boldsymbol{a}_1, \boldsymbol{a}_1)/(\boldsymbol{a}_i, \boldsymbol{a}_1), \ i=2, 3, \cdots, n. \tag{58}$$

在这一步骤的最后，就得到了向量系

$$\boldsymbol{a}_1^{(1)}, \boldsymbol{a}_2^{(1)}, \cdots, \boldsymbol{a}_n^{(1)},$$

后面 $n-1$ 个向量都和第一个向量直交．

第二步：对 $i=3, 4, \cdots, n$，在 $\boldsymbol{a}_2^{(1)}$ 和 $\boldsymbol{a}_i^{(1)}$ 规定的平面上，确定向量 $\boldsymbol{a}_i^{(2)}$，使它和 $\boldsymbol{a}_2^{(1)}$ 直交，并令 $\boldsymbol{a}_1^{(2)}=\boldsymbol{a}_1^{(1)}$，$\boldsymbol{a}_2^{(2)}=\boldsymbol{a}_2^{(1)}$．确定 $\boldsymbol{a}_i^{(2)}$ 的方法和第一步中所用的完全一样．

这一步实现以后，就得到了向量系

$$\boldsymbol{a}_1^{(2)},\ \boldsymbol{a}_2^{(2)},\ \cdots,\ \boldsymbol{a}_n^{(2)}.$$

以上的步骤进行 $n-1$ 次，最后得到的向量系

$$\boldsymbol{a}_1^{(n-1)},\ \boldsymbol{a}_2^{(n-1)},\ \cdots,\ \boldsymbol{a}_n^{(n-1)}$$

便是我们所需要的直交系 $\boldsymbol{v}_1,\ \boldsymbol{v}_2,\ \cdots,\ \boldsymbol{v}_n$.

为了保证序列 $\|\boldsymbol{v}_1\|,\ \|\boldsymbol{v}_2\|,\ \cdots,\ \|\boldsymbol{v}_n\|$ 为单调不增，在第 i 步进行直交化以后，得到的新的向量为

$$\boldsymbol{a}_{i+1}^{(i)},\ \boldsymbol{a}_{i+2}^{(i)},\ \cdots,\ \boldsymbol{a}_n^{(i)}.$$

选上列向量中的长度最大者移到 $\boldsymbol{a}_{i+1}^{(i)}$ 的位置，并记之为 $\boldsymbol{a}_{i+1}^{(i+1)}$. 这个向量就是 $\boldsymbol{v}_{i+1}$. 然后再进行第 $i+1$ 步.

为了说明上述直交化方法的数值稳定性的不同，今举一例.

例：设

$$\boldsymbol{a}_1=\begin{bmatrix}1\\1+\varepsilon\\1\\1\end{bmatrix},\quad \boldsymbol{a}_2=\begin{bmatrix}1\\1\\1+\varepsilon\\1\end{bmatrix},\quad \boldsymbol{a}_3=\begin{bmatrix}1\\1\\1\\1+\varepsilon\end{bmatrix},$$

其中，$0<|\varepsilon|\ll 1$，使得在计算机上 $1+\varepsilon$ 可以准确表示，而数 $1+2\varepsilon+\varepsilon^2$ 就被“舍入”为 $1+2\varepsilon$.

用古典的 Gram-Schmidt 直交化方法，得到的所谓直交系是：

$$\boldsymbol{v}_1=\begin{bmatrix}1\\1+\varepsilon\\1\\1\end{bmatrix},\quad \boldsymbol{v}_2=\begin{bmatrix}0\\-\varepsilon\\\varepsilon\\0\end{bmatrix},\quad \boldsymbol{v}_3=\begin{bmatrix}0\\-\varepsilon\\0\\\varepsilon\end{bmatrix}.$$

注意到向量 $\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$ 之间夹角 θ 的余弦为

$$\cos\theta=\frac{(\boldsymbol{v}_2,\ \boldsymbol{v}_3)}{\|\boldsymbol{v}_2\|\|\boldsymbol{v}_3\|}=\frac{1}{2},$$

所以 $\theta=60^\circ$，从而可以看出向量 $\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$ 远远并不直交！

另一方面，用改进的 Gram-Schmidt 直交化方法得到的向量系是：

$$\boldsymbol{v}_1=\begin{bmatrix}1\\1+\varepsilon\\1\\1\end{bmatrix},\quad \boldsymbol{v}_2=\begin{bmatrix}0\\-\varepsilon\\\varepsilon\\0\end{bmatrix},\quad \boldsymbol{v}_3=\begin{bmatrix}0\\-\frac{1}{2}\varepsilon\\-\frac{1}{2}\varepsilon\\0\end{bmatrix},$$

这时，$(\boldsymbol{v}_2, \boldsymbol{v}_3)=0$，这就是说 $\boldsymbol{v}_2$ 和 $\boldsymbol{v}_3$ 的确为直交.

这两种直交化方法，从“理论”上讲是等价的，而在计算机上进行计算的结果却可以有很大差异．这是由于两种算法的计算过程不同，舍入误差的传播也不同，这是算法的数值稳性程度不同的表现．这种现象，在计算数学中带有普遍性，它告诉我们，切勿把数学上等价的表达式随意应用!

3-4 LDL^T 分解法

若向量系 $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ 为线性独立，则对任何 k，$1\leqslant k\leqslant n$，矩阵 $A_k^TA_k$ 总为对称正定，从而存在有唯一的分解式：

$$A_k^TA_k=L_kD_kL_k^T, \tag{59}$$

其中的 L_k 是一个 $k\times k$ 阶的单位下三角阵，D_k 是一个 $k\times k$ 阶的对角阵

$$D_k=\mathrm{diag}(d_1, d_2, \cdots, d_k),$$
$$d_i>0,\ i=1, 2, \cdots, k.$$

公式(59)就是通常所说的对称正定矩阵的 Cholesky 分解，并有有效的算法来实现这种分解.

因为矩阵乘积的行列式等于各因子矩阵行列式之积，于是有

$$\begin{aligned}G[A_k]&=\det(L_k)\cdot\det(D_k)\cdot\det(L_k^T)\\&=\det(D_k)=d_1d_2\cdots d_k,\ k=1, 2, \cdots, n.\end{aligned}$$

在进行矩阵分解时，为加强算法的数值稳定性，常采用选主元的方法，这样就可以保证序列 $d_1, d_2, \cdots, d_n$ 为单调不增．与 **3-2** 中的 QR 分解公式相比较，可知

$$L_k D_k L_k^T = \widetilde{R}_k^T \widetilde{R}_k,\ k=1,\ 2,\ \cdots,\ n.$$

从而可知

$$|\rho_{ii}| = \sqrt{d_i},\ i=1,\ 2,\ \cdots,\ k. \tag{60}$$

据此便得到了向量 $\boldsymbol{a}_k$ 关于向量系的相对相关性指标：

$$E[\boldsymbol{a}_k | \boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}] = \sqrt{d_k} / \sqrt{d_1},$$

若向量系 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n$ 已规范化，则 $d_1=1$.

§4 矩阵的奇异度、条件数和伪秩

4-1 引言

在本章一开始我们就谈到方阵的奇异性问题．对方阵求逆而言，奇异性意味着间断性．现在我们把奇异性概念推广到一般的 $m\times n$ 阶矩阵上来．

定义 5 设 A 是一个 $m\times n$ 阶矩阵，若 A 为亏秩，则称之为**奇异**．若 A 为满秩，称之为**正则**．

根据第 3 章的讨论，我们已经知道，对于求一般 $m\times n$ 阶矩阵 A 的广义逆 A^+ 来说，奇异性也同样意味着间断性．但是和方阵在通常意义下的求逆问题而言，却有根本的不同．其一是，亏秩矩阵 A 的广义逆 A^+ 总是存在的而且是唯一的．其二是，亏秩矩阵的广义逆虽然在理论上是不连续的，但是却可以用保秩变形的方法来消除这种间断性使之变为连续．这一点十分重要，它为使用广义逆以克服在通常意义下求逆时不可克服的奇异性所带来的严重困难，奠定了理论基础．矩阵的奇异性或亏秩，是一个定性的概念，为了适应数值计算的需

要，有必要使之定量化．目前，有些人常说一个矩阵是“接近奇异的”，或“几乎奇异的”，这是不确切的．从理论上讲，矩阵的奇异性或亏秩，是矩阵的列向量（对于 $m\times n$ 阶矩阵，我们假定 $m\geq n$！）之间的线性相关性直接相联系的，据此，我们可以把对于线性相关性的定量讨论应用到对奇异性的定量讨论上来．为此目的，我们首先给出以下的定义．

定义 6 我们假定 $m\times n$ 阶矩阵 $A=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$ 中列向量的顺序，使逐次的相对相关性指标为单调不增．设 ε 为给定的一个正的小数，若 A 为满秩，但

$$E[\boldsymbol{a}_{k+1}|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant\varepsilon,$$

我们就称矩阵 A 具有 **$n-k$ 阶的 ε-奇异度**，其中若 $k=n-1$，这时矩阵 A 具有 1 阶的 ε-奇异度．

定义 7 若

$$E[\boldsymbol{a}_{k+1}, \boldsymbol{a}_1, \cdots, \boldsymbol{a}_k]\leqslant\varepsilon,\ E[\boldsymbol{a}_k|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}]>\varepsilon,$$

我们就称矩阵 A 的 **ε-秩**为 k．

在这里我们顺便指出，从实际计算的角度来讲，矩阵的**真秩**，往往是难以确定的．而实际上，在许多算法中，我们并不需要矩阵的真秩，重要的是，为使算法能得到较好的结果，应当设法确定它的某一个 ε-秩，关于 ε 的选定是极其重要的．例如，若计算有关 Hilbert 矩阵 H_n 的问题，在 n 比较大时，如果把它作为满秩来处理，得到的计算结果将不堪设想！而众所周知，对任何 n，H_n 总是正定的，从而也是满秩的．一般说来，奇异性并不可怕，困难在于问题的病态性质．

4-2 矩阵的 ε-秩和矩阵的拟条件数

前面我们已经给出了矩阵的条件数概念，并证明矩阵的谱条件数可以表示成该矩阵的最大和最小正奇异值之比．但是，计算矩阵的奇异值需要较大的工作量，而且，一般说来，这

种估计往往偏于保守．这里我们将对于矩阵的条件数给出一种简便易行的估计方法．为了这一目的，我们先来考虑矩阵的条件数和相关性指标之间的关系．

设矩阵 $A=[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n]$ 已经规范化，并假定诸列向量的顺序可以保证各次的相对相关性指标为单调不增．我们先设 $\mathrm{rank}(A)=n$，对 A 作 QR 分解得

$$A=QR \tag{61}$$

其中的 Q 是一个 $m\times n$ 阶列直交矩阵，R 是一个上三角阵，其主对角元为 $\rho_{11}, \rho_{22}, \cdots, \rho_{nn}$，并且

$$1=|\rho_{11}|\geqslant|\rho_{22}|\geqslant\cdots\geqslant|\rho_{nn}|. \tag{62}$$

容易证明矩阵 R 和 A 有共同的奇异值．记其最大和最小奇异值分别为 λ_1 和 λ_n．因为 $R^TR=A^TA$，所以 A 和 R 有共同的奇异值．此外，由于 R 和 R^T 有共同的非零奇异值，并且

$$\boldsymbol{e}_1^TR^TR\boldsymbol{e}_1=\rho_{11}^2,\ \boldsymbol{e}_n^TRR^T\boldsymbol{e}_n=\rho_{nn}^2,$$

根据特征值的极性可知

$$\lambda_1\geqslant|\rho_{11}|,\ \lambda_n\leqslant|\rho_{nn}|. \tag{63}$$

于是便有

定理 5 在前述假设下，必有

$$k(A)=\frac{\lambda_1}{\lambda_n}\geqslant\frac{|\rho_{11}|}{|\rho_{nn}|}=\{E[\boldsymbol{a}_n|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{n-1}]\}^{-1}. \tag{64}$$

因此，若

$$E[\boldsymbol{a}_n|\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{n-1}]=\frac{|\rho_{nn}|}{|\rho_{11}|}=|\rho_{nn}|\leqslant\varepsilon,$$

则

$$k(A)\geqslant\frac{1}{\varepsilon}.$$

这就是说，$|\rho_{nn}|$ 愈小，谱条件数 $k(A)$ 就愈大．另一方面，

$$\sum_{i=1}^n\lambda_i^2=\mathrm{tr}(A^TA)=\sum_{i=1}^n\|\boldsymbol{a}_i\|^2\leqslant n,$$

据此可知$\lambda_1<\sqrt{n}$. 这就是说，最大奇异值有一个上界为$\sqrt{n}$. 所以，在规范化的情况下，条件数$k(A)$的大小主要决定于最小奇异值λ_n.

若矩阵A为非奇异，则ρ_{nn}和λ_n都异于零；若矩阵A为奇异，则二者同时为零. 由于ρ_{nn}和λ_n都是矩阵元素的连续函数，所以有：

1° 当$\rho_{nn}\to 0$时，必有$\lambda_n\to 0$；反之亦然.

此外，若$\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n$为直交系，则因

$$A^TA=\mathrm{diag}(\rho_{11}^2, \cdots, \rho_{nn}^2),$$

从而可知：

2° $\lambda_n=|\rho_{nn}|$.

一般说来，$|\rho_{nn}|\geq\lambda_n$，它仅为λ_n的一种近似. 从实际的计算结果可以看出，λ_n愈小，$|\rho_{nn}|$就愈和λ_n近似.

定义 8 设$A=[\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n]$为秩是$k$的$m\times n$阶矩阵，在本节的假设下，今规定

$$k_c(A)=\{E[\boldsymbol{\alpha}_k|\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{k-1}]\}^{-1} \tag{65}$$

为矩阵A的**拟条件数**.

上述拟条件数的定义是对在理论上秩为k的矩阵而言的. 但是，在许多实际计算问题中，所遇到的矩阵A常常是满秩的(有时甚至在理论上就可以证明矩阵是满秩的，如Hilbert矩阵)，如果我们按满秩进行处理，不仅得不到好的计算结果，甚至会使算法不能进行下去. 这时，一种常用的方法，就是降秩. 从而产生了伪秩(pseudo rank)这个概念. 顾名思义，所谓伪秩，就不是矩阵的真秩(即理论上的秩). 前面我们提的ε-秩这个概念，其目的也就在于此. 现在我们假设对于选定的正数ε，已知矩阵A的ε-秩为k，并对这种情形来考虑问

题的条件．对于矩阵 A 的假设仍和前面一样，并对 A 作QU分解：

$$A=QU, \tag{66}$$

这时，Q 为一个 $m\times n$ 阶列直交矩阵，U 是一个 $n\times n$ 阶上梯形矩阵，其具体形式如下：

$$U=\begin{bmatrix} R & U_1 \\ 0 & U_2 \end{bmatrix}, \tag{67}$$

其中的 R 是一个 $k\times k$ 阶的非奇异上三角阵，它的主对角元 $|\rho_{11}|$，$|\rho_{22}|$，$\cdots$，$|\rho_{kk}|$ 系单调不增，它们就是相应的相对相关性指标（注意，对规范化的矩阵，$|\rho_{11}|=1$）．

定义 9 我们规定

$$\{E[\boldsymbol{a}_k|\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_{k-1}]\}^{-1}=\frac{1}{|\rho_{kk}|} \tag{68}$$

为确定矩阵 A 的 ε-秩为 k 时的拟条件数，并记之为 $k_q(A_\varepsilon)$．

为使读者对这样定义的拟条件数的意义有进一步的了解，今讨论于下．这时，矩阵 A 的分解式(66)可以表示成

$$A=QU=A_1+\delta A_1, \tag{69}$$

其中，

$$A_1=Q\begin{bmatrix} R & U_1 \\ 0 & 0 \end{bmatrix},\quad \delta A_1=Q\begin{bmatrix} 0 & 0 \\ 0 & U_2 \end{bmatrix}. \tag{70}$$

显然矩阵 A_1 的秩为 k，而矩阵 A 的真秩则可能大于 k 甚至为满秩．因为 $\mathrm{rank}(A_1)=k$，所以 A_1 的后 $n-k$ 个奇异值均为零．设矩阵 A 的后 $n-k$ 个奇异值为

$$\lambda_{k+1},\ \cdots,\ \lambda_n,$$

则根据奇异值摄动定理有

$$\lambda_j\leqslant\|\delta A_1\|=\|U_2\|,\ j=k+1,\ \cdots,\ n. \tag{71}$$

因为矩阵 A 的 ε-秩为 k，所以有

$$E[\boldsymbol{a}_j|\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_k]<\varepsilon,\ j=k+1,\ \cdots,\ n.$$

从而可知

$$\|U_2\| < \sqrt{n-k}\ \varepsilon. \tag{72}$$

于是便得到了

定理 6 设矩阵 A 的 ε-秩为 k，则恒有关系

$$\lambda_j < \sqrt{n-k}\ \varepsilon, \quad j=k+1, \cdots, n. \tag{73}$$

这就说明，若相对相关性指标 $|\rho_{k+1,k+1}|$ 很小的时候，矩阵 A 有 $n-k$ 个小奇异值. 就极限情形来说，当 $|\rho_{k+1,k+1}| \to 0$ 时，矩阵 A 有 $n-k$ 个奇异值 $\lambda_j \to 0$, $j=k+1, \cdots, n$.

在实际计算时，为了控制矩阵 A 病态的危害程度，常根据适当选取的正数 ε 来确定 A 的一个降秩近似，如前面矩阵 A_1. 在算法中采取了这种措施，就可以避免舍入误差影响的恶化，从而提高算法的数值稳定性。

第 5 章

广义逆矩阵和线性最小二乘问题的计算方法

§1 计算广义逆矩阵的基本原则

设 A 是一个 $m\times n$ 阶实矩阵，除非特别说明，我们总假定 $m\geqslant n$. 计算广义逆矩阵 A^+ 的一个基本原则是利用矩阵 A 的满秩分解

$$A=BC, \tag{1}$$

其中的 B 和 C 分别为 $m\times r$ 阶列满秩和 $r\times n$ 阶行满秩矩阵，而 $r=\operatorname{rank}(A)$. 这时 A 的广义逆矩阵 A^+ 便可以表示成

$$A^+=C^+B^+. \tag{2}$$

因为 B 和 C 都是满秩矩阵，所以它们的广义逆矩阵 B^+ 和 C^+ 便可以分别表示为:

$$B^+=(B^TB)^{-1}B^T,\ C^+=C^T(CC^T)^{-1}, \tag{3}$$

其中的 B^TB 和 CC^T 都是 $r\times r$ 阶对称正定方阵，于是广义逆矩阵 A^+ 便可以表示成:

$$A^+=C^T(CC^T)^{-1}(B^TB)^{-1}B^T. \tag{4}$$

这时，超定线性方程组

$$A\boldsymbol{x}=\boldsymbol{b} \tag{5}$$

的极小最小二乘解 $\boldsymbol{x}$ 就可以用公式

$$\boldsymbol{x}=A^+\boldsymbol{b}=C^T(CC^T)^{-1}(B^TB)^{-1}B^T\boldsymbol{b} \tag{6}$$

表示出来，如果矩阵 A 为列满秩，即 $\operatorname{rank}(A)=n$，则公式 (4) 就简化为

$$A^+=(A^TA)^{-1}A^T, \tag{7}$$

而相应的极小最小二乘解 $\boldsymbol{x}$ 的表达式 (6) 就化为

$$\boldsymbol{x}=(A^TA)^{-1}A^T\boldsymbol{b}. \tag{8}$$

上面计算 $A^\dagger$ 的公式 (4) 以及计算极小最小二乘解的公式 (6) 中都有两次矩阵求逆，这仅仅是形式上的，在实际计算时还可以大为简化. 例如，令 $\boldsymbol{u}=B^T\boldsymbol{b}$，解方程

$$(B^TB)(CC^T)\boldsymbol{v}=\boldsymbol{u} \tag{9}$$

求出 $\boldsymbol{v}$，于是就有 $\boldsymbol{x}=C^T\boldsymbol{v}$. 而在计算 $\boldsymbol{v}$ 时，只需要作一次矩阵分解就行了，并不需要计算两次逆矩阵.

本章将介绍几种常用的关于矩阵 A 的满秩分解方法，并利用这些方法来计算广义逆矩阵 A^+ 以及相应的超定线性方程组的极小最小二乘解 $\boldsymbol{x}$.

和一般的计算问题一样，在考虑计算方法时，应当考虑下列三个方面的问题:

(i) 计算工作量的大小问题;

(ii) 存贮方法和存贮量的大小问题;

(iii) 算法的数值稳定性问题.

在本章中我们介绍有关算法时，将分别予以说明.

§2 Gauss 消去法

2-1 Gauss 消去法的基本步骤

Gauss 消去法实际上就是矩阵的 LU 分解法，是众所周知的解线性方程组的一种常用的方法. 这种方法也可以用来对矩阵 A 作满秩分解，它具有计算量小的优点. 为了说明这个方法的基本思想和步骤，我们先把矩阵 A 用它的 $m\times n$ 个元素 a_{ij} 表示成通常的形式:

$$
A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{10}
$$

并假定 $\mathrm{rank}(A)=r$.

消去法的过程是用一种特殊形式的矩阵依次左乘矩阵 A，逐步使矩阵 A 中位于主对角元下方的元素都化为零. 为了使算法不遭到不应有的终断并且加强算法的数值稳定性，需要采用选主元的措施. 每一列消元为算法的一个“大步”，因为矩阵 A 的秩为 r，所以经过 r 个“大步”消元，矩阵 A 就化为一个上梯形阵

$$
\begin{bmatrix} \diagdown \; U \\ \hline O \end{bmatrix},
$$

U 是一个 $r\times n$ 阶上梯形阵，其下方的元素都是零.

消元的步骤如下：

1° 选主元　为了提高算法的数值稳定性，以就总体选主元为宜. 这就是说，在执行第一大步的开始，先从矩阵 A 的全体元素中选取绝对值最大的元素，设其为 $a_{p,q}$，若存在有几个这样的元素，可取进行元素比较时第一次发现的那一个，并把它经过行、列交换移到 A 的左上角的位置作为第一大步的主元，这相当于对矩阵 A 分别左乘一个排列阵以使 A 的 1, p 两行交换，再右乘一排列阵，以使新矩阵的 1, q 两列交换. 记最后所得矩阵为 $A^{(1)}=[a_{ij}^{(1)}]$.

2° 消元　记

$$
l_{i1}=\frac{a_{i1}^{(1)}}{a_{11}^{(1)}},\ i=2,\ 3,\ \cdots\cdots,\ m,
$$

依次乘 $A^{(1)}$ 的第一行并把所得结果从第 i 行中减去，于是，矩阵 $A^{(1)}$ 中第一列主对角元以下的元素便都化为零. 这一

大步的全过程，相当于用矩阵

$$F_1=\begin{bmatrix} 1 & & & \\ -l_{21} & 1 & 0 & \\ -l_{31} & & \ddots & \\ \vdots & 0 & & \ddots \\ -l_{m1} & \cdots\cdots & & 1 \end{bmatrix} \tag{11}$$

左乘矩阵 $A^{(1)}$. 结果为

$$F_1A^{(1)}=\begin{bmatrix} \times & \times\cdots\cdots\times \\ 0 & \times\cdots\cdots\times \\ 0 & \times\cdots\cdots\times \\ \cdots & \cdots\cdots\cdots\cdots \\ 0 & \times\cdots\cdots\times \end{bmatrix}. \tag{12}$$

在进行第二大步的开始，先从矩阵 (12) 中位于第一行之下而位于第一列之右的那个 $(m-1)\times(n-1)$ 阶子矩阵中，选取绝对值最大的元素把它移到 (2, 2) 的位置，作为主元. (12) 右方的矩阵在根据选主元的要求交换行列后，所得矩阵为 $A^{(2)}=[a_{ij}^{(2)}]$，令

$$l_{i2}=\frac{a_{i2}^{(2)}}{a_{ii}^{(2)}},\ i=3,\ 4,\ \cdots,\ m. \tag{13}$$

仿前，构造矩阵

$$F_2=\begin{bmatrix} 1 & & & & \\ 0 & 1 & & 0 & \\ 0 & -l_{32} & 1 & & \\ \vdots & \vdots & & \ddots & \\ \vdots & \vdots & 0 & & \ddots \\ 0 & -l_{m2} & & & 1 \end{bmatrix}, \tag{14}$$

并用它左乘 $A^{(2)}$. 这就完成了第二大步.

如此进行下去，因为 A 的秩为 r，左乘 F_i 不会改变它的

秩，在进行了 r 个大步以后，原矩阵 A 就化为一个上梯形阵了．整个过程，包括选主元和消元，可以表示为

$$F_rF_{r-1}\cdots\cdots F_1P_1AP_2=\begin{bmatrix}U\\0\end{bmatrix}, \tag{15}$$

其中的 P_1, P_2 为两个排列阵，F_1, F_2, ……, F_r 为各次消元时所用的 F 阵（有时称之为 Frobenius 矩阵），U 是一个 $r\times n$ 阶上梯形阵．这 r 个 F 阵的乘积是一个单位下三角阵，它的具体形式为

$$F_rF_{r-1}\cdots\cdots F_1=\begin{bmatrix}1 & & & & & & & \\ -l_{21} & 1 & & & & \mathbf{0} & & \\ -l_{31} & -l_{32} & 1 & & & & & \\ \vdots & & \ddots & \ddots & & & & \\ -l_{r+1,1} & \cdots\cdots & & -l_{r+1,r} & 1 & & & \\ \vdots & & & \vdots & 0 & \ddots & & \\ \vdots & & & \vdots & \vdots & & \ddots & \\ -l_{m1} & \cdots\cdots & & -l_{mr} & 0 & \cdots & 0 & 1\end{bmatrix} \tag{16}$$

它的逆矩阵是

$$(F_rF_{r-1}\cdots\cdots F_1)^{-1}=\begin{bmatrix}1 & & & & & & \\ l_{21} & 1 & & & \mathbf{0} & & \\ \vdots & \ddots & \ddots & & & & \\ & & & 1 & & & \\ l_{r+1,1} & \cdots\cdots & & l_{r+1,r} & 1 & & \\ \vdots & & & & 0 & \ddots & \\ l_{m,1} & \cdots\cdots & & l_{mr} & 0 & \cdots\ 0 & 1\end{bmatrix}=L_1（设） \tag{17}$$

于是便得到最后的矩阵分解

$$P_1AP_2=L_1\begin{bmatrix}U\\0\end{bmatrix}. \tag{18}$$

考虑到 (18) 右端 U 下方的矩阵是 $(m-r)\times n$ 阶零矩阵，在

运算时，矩阵 L_1 的后 $m-r$ 列实际上不起作用，记其前 r 列为 L，则 (18) 可以表示成

$$P_1AP_2=LU, \tag{19}$$

这里的 L 是一个 $m\times r$ 阶的单位下梯形阵，U 是一个 $r\times n$ 阶的上梯形阵，它们的秩都是 r。所以在消元最后，我们便得到了矩阵 P_1AP_2 的满秩分解。

最后我们指出，在选主元过程中进行矩阵的行、列交换时，应注意以下两点：

(1) 左乘排列阵进行行交换时，方程 $A\boldsymbol{x}=\boldsymbol{b}$ 右端向量 $\boldsymbol{b}$ 的相应分量也必须同时交换，而 $\boldsymbol{x}$ 的分量保持原来的次序。

(2) 右乘排列阵进行列交换时，$\boldsymbol{x}$ 的相应分量应同时进行交换，而右端向量 $\boldsymbol{b}$ 的分量次序不变。

2-2 Gauss 消去法的实现

1° 矩阵 L 和 U 的元素计算公式。 如果不考虑矩阵行列交换对矩阵元素下标所带来的影响，矩阵 L 和 U 的元素可依次按下列公式计算：

对 $p, q=1, 2, \cdots\cdots, r$ 依次计算

$$u_{pj}=a_{pj}-\sum_{k=1}^{p-1}l_{pk}u_{kj},\ j=p,\ p+1,\ \cdots\cdots,\ n; \tag{20}$$

$$l_{iq}=\left(a_{iq}-\sum_{k=1}^{p-1}l_{ik}u_{kq}\right)/u_{qq},\ i=q+1,\ \cdots\cdots,\ m. \tag{21}$$

为了提高算法的数值稳定性，在计算内积累加时应采用双精度运算。

2° 关于矩阵 U 和 L 元素的存放。 当元素 u_{pj} 和 l_{iq} 依次计算出来以后，随即把它们存放到元素 a_{pj} 和 a_{iq} 的位置上。这就是说，把原来存放矩阵 A 的元素的单元，逐次换成 U 和 L 的元素：

$$
\begin{bmatrix}
u_{11} & u_{12} \cdots\cdots\cdots & u_{1n} \\
l_{21} & u_{22} \cdots\cdots\cdots & u_{2n} \\
& \cdots\cdots\cdots\cdots\cdots\cdots & \\
l_{r1} & l_{r2} \cdots u_{rr} \cdots & u_{rn}
\end{bmatrix}. \tag{22}
$$

3° 矩阵伪秩的确定． 矩阵 A 的秩实际上事先并不知道，我们只能在计算过程中确定它的一个适当的伪秩，其法如下：

给定控制量 $\varepsilon>0$，（一般比计算机的相对精度要大一些），如果在第 $r+1$ 大步所选主元的绝对值小于 ε 时，就确定 r 为矩阵 A 的 ε-秩．

4° 计算极小最小二乘解时加强算法数值稳定性的方法．

根据前述矩阵的满秩 LU 分解 (19)，计算方程 $A\boldsymbol{x}=\boldsymbol{b}$ 的极小最小二乘解 $\boldsymbol{x}$ 的基本公式为：

$$
\begin{aligned}
\boldsymbol{x} &= U^T(UU^T)^{-1}(L^TL)^{-1}L^T\boldsymbol{b} \\
&= U^T(L^TLUU^T)^{-1}L^T\boldsymbol{b}.
\end{aligned} \tag{23}
$$

为了改善算法的数值稳定性，可令

$$
U = D\hat{U}, \tag{24}
$$

其中 D 为对角阵

$$
D=\mathrm{diag}(u_{11}, \cdots\cdots, u_{rr}). \tag{25}
$$

矩阵 $\hat{U}$ 是一个 $r\times n$ 阶单位上梯形阵，它的元素 $\hat{u}_{ij}$ 的绝对值不大于 1，

$$
|\hat{u}_{ij}|\leqslant 1, \tag{26}
$$

这有利于控制舍入误差的传播．实际上我们使用的不是 LU 分解，而是 A 的 $\mathrm{LD\hat{U}}$ 分解．这时，公式 (23) 就化为

$$
\boldsymbol{x}=\hat{U}^T(L^TLD\hat{U}\hat{U}^T)^{-1}L^T\boldsymbol{b}. \tag{27}
$$

5° 矩阵 A 为列满秩的情形．

这时，因为 $\mathrm{rank}(A)=n$，(23) 中的 U 为 $n\times n$ 阶非奇异

方阵，故有

$$U^T(UU^T)^{-1}=U^{-1},$$

从而 (23) 就简化为

$$\boldsymbol{x}=U^{-1}(L^TL)^{-1}L^T\boldsymbol{b}. \tag{28}$$

这个计算公式具有较好的数值稳定性. 因为选主元常使 L 为良态，从而矩阵 A 的病态性质决定于矩阵 U. 原来矩阵 UU^T 求逆的条件数为

$$k(UU^T)=k^2(U), \tag{29}$$

由于问题现在化为 U 的求逆，所以问题 (28) 便具有较好的条件.

2-3 运算量和误差分析

如果把一次加法和一次乘法合在一起算作一次运算，并且 $m, n, r\gg1$，则对矩阵 A 进行 LU 分解所需要的总运算次数的阶为

$$r\left[mn-\frac{1}{2}(m+n)r+\frac{1}{3}r^2\right]. \tag{30}$$

利用 Wilkinson 向后误差分析的方法可以证明，用 Gauss 消去法所得到的 LU 分解，是某一摄动矩阵 $A+\delta A$ 的精确分解，

$$LU=A+\delta A. \tag{31}$$

如果使用按列选主元的方法，则对 δA 可得出下列估计：

$$\left.\begin{array}{l}\delta a_{1j}=0,\ j=1,2,\cdots,n.\\ |\delta a_{ij}|\leqslant|l_{ij}u_{jj}|\eta,\ j=1,\cdots,r;\ i=j+1,\cdots,m,\\ |\delta a_{ij}|\leqslant|u_{ij}|\eta,\ i=2,\cdots,r;\ j=i,\cdots,n,\\ |\delta a_{ij}|\leqslant\varepsilon,\ i=r+1,\cdots,m;\ j=r+1,\cdots,n,\end{array}\right\} \tag{32}$$

其中的 η 表示计算机的精度.

在这里我们不对误差分析作具体推导，而只列出结果，因

为我们的目的仅在于对本章所介绍的某些算法作对比. 对误差分析有兴趣的读者可参阅 Wilkinson 的专著或本章所列出的有关文献.

§3 QR 或 QU 分解法

3-1 Householder 变换

为了建立算法的需要，我们先来介绍一种结构简单的对称直交变换. 设 $\boldsymbol{u}$ 为 $\mathscr{R}^m$ 中的一个长度为 1 的向量. 令

$$H=I+\beta^{-1}\boldsymbol{u}\boldsymbol{u}^T, \tag{33}$$

矩阵 H 显然有对称性. 欲其为一直交阵, 必须且只须

$$HH^T=(I+\beta^{-1}\boldsymbol{u}\boldsymbol{u}^T)(I+\beta^{-1}\boldsymbol{u}\boldsymbol{u}^T)^T=I,$$

据此可推得

$$\beta^{-1}=-2.$$

代入 (33) 中便得到对称直交阵

$$H=I-2\boldsymbol{u}\boldsymbol{u}^T. \tag{34}$$

直交阵 (34) 便是著名的 Householder 矩阵.

现在我们来进一步考虑, 对给定的长度相等的两个 m 维向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$, 如何选择向量 $\boldsymbol{u}$ 使所得的 Householder 矩阵 H 具有性质:

$$H\boldsymbol{x}=\boldsymbol{y}. \tag{35}$$

把 (34) 中的 H 表达式代入 (35) 中, 便得到关系

$$2(\boldsymbol{u}^T\boldsymbol{x})\boldsymbol{u}=\boldsymbol{x}-\boldsymbol{y}.$$

这就是说, 向量 $\boldsymbol{u}$ 应与向量 $\boldsymbol{x}-\boldsymbol{y}$ 平行. 由于 H 表达式中的 $\boldsymbol{u}$ 为单位长, 所以可令

$$\boldsymbol{u}=\frac{\boldsymbol{x}-\boldsymbol{y}}{\|\boldsymbol{x}-\boldsymbol{y}\|}, \tag{36}$$

这里应假定 $\boldsymbol{x}\neq\boldsymbol{y}$. 于是便得到了我们所需要的 H 阵:

$$H=I-2\frac{(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})^{T}}{\|\boldsymbol{x}-\boldsymbol{y}\|^{2}}. \tag{37}$$

现在我们来确定向量 $\boldsymbol{u}$, 使矩阵 H 具有性质:

$$H\boldsymbol{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_{p-1} \\ -\sigma\xi_p \\ x_{p+1} \\ \vdots \\ x_{l-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}=\boldsymbol{y}, \quad \sigma=\pm 1. \tag{38}$$

因为向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 的长度相等, 首先可知

$$\xi_p=(x_p^2+\sum_{i=l}^{m}x_i^2)^{1/2}.$$

这里的 p 和 l 是事先给定的非负整数. 这就是说, 矩阵 H 使 $\boldsymbol{x}$ 的后 $m-l+1$ 个分量化为零, 保持前面的分量除第 p 个外都不变. 于是, $\boldsymbol{y}$ 的第 p 个分量

$$-\sigma\xi_p,$$

仅绝对值被固定, 符号可根据需要来选择. 为了使计算 u_p 时避免减法以改善稳定性, 可令

$$\sigma=\begin{cases} +1, & \text{若 } x_p\geqslant 0, \\ -1, & \text{若 } x_p<0. \end{cases} \tag{39}$$

关于向量 $\boldsymbol{u}$ 的分量 u_1, ……, u_m, 可计算如下:

令
$$s=-\sigma(x_p^2+\sum_{i=l}^{m}x_i^2)^{1/2},$$
$$\begin{aligned}&u_i=0,\ i=1,\ \cdots\cdots,\ p-1,\\&u_p=x_p-s,\\&u_i=0,\ i=p+1,\ \cdots\cdots,\ l-1,\\&u_i=x_i,\ i=l,\ \cdots\cdots,\ m.\end{aligned}\tag{40}$$
令
$$\beta=su_p,\tag{41}$$
然后便得到所需要的 H 阵
$$\begin{aligned}&H=I+\beta^{-1}\boldsymbol{u}\boldsymbol{u}^T,\ 若\ \beta\neq0,\\&H=I,\qquad\qquad 若\ \beta=0.\end{aligned}\tag{42}$$

3-2 矩阵的 QR 和 QU 分解

有了 Householder 矩阵，我们就可以对 $m\times n$ 阶矩阵 $A=[\boldsymbol{a}_1,\ \cdots,\ \boldsymbol{a}_n]$ 进行满秩分解。和 Gauss 消去法相类似，若 $\text{rank}(A)=r\leqslant n$，我们也分 r 步来进行。

第一步，确定 H 阵 H_1，即，确定单位向量 $\boldsymbol{u}_1$，它所构造的 Householder 矩阵 H_1，使
$$H_1\boldsymbol{a}_1=-\sigma\|\boldsymbol{a}_1\|\boldsymbol{e}_1,$$
其中的 $\boldsymbol{e}_1$ 是 m 维空间的第一个单位坐标向量，$\sigma=\pm1$，若 $\boldsymbol{a}_1$ 的第一个分量非负，取 $\sigma=+1$，否则，取 $\sigma=-1$. 在 (40) 中分别令 $p=1$, $l=2$，便可依次算出
$$s=-\sigma\|\boldsymbol{a}_1\|,$$
$$u_1=a_{11}-s,\ u_i=a_{1i},\ i=2,\ \cdots\cdots,\ m,$$
$$\beta=su_1,$$
然后按 (42) 即可算出 H_1，以其左乘矩阵 A 得
$$H_1A=[H_1\boldsymbol{a}_1,\ \cdots\cdots,\ H_1\boldsymbol{a}_n]=A^{(1)}(设)$$
从而就完成了第一步消元.

第二步：把矩阵 $A^{(1)}$ 右下角那个 $(m-1)\times(n-1)$ 阶子

矩阵记作 $\hat{A}^{(1)}$，

$$\hat{A}^{(1)}=[\hat{\boldsymbol{a}}_2,\ \cdots\cdots,\ \hat{\boldsymbol{a}}_n],$$

它是一个 $(m-1)\times(n-1)$ 阶矩阵. 用和第一步完全相似的方法确定 $(m-1)\times(m-1)$ 阶的 H 阵 $\hat{H}_2$，使 $\hat{H}_2\hat{\boldsymbol{a}}_2$ 的分量除第一个外都是零. 令

$$H_2=\begin{bmatrix}1 & 0\\ 0 & \hat{H}_2\end{bmatrix}.$$

计算 $H_2A^{(1)}$，并记其结果为 $A^{(2)}$. 这样就完成了第二步. 在第二步消元时，矩阵 $A^{(1)}$ 的第一行和第一列都不发生变化.

记 $A^{(2)}$ 右下角那个 $(m-2)\times(n-2)$ 阶子矩阵为 $\hat{A}^{(2)}$，并仿前进行第三步.

这样继续进行下去，到第 r 步完成，矩阵 A 就化为一个上梯形阵

$$H_rH_{r-1}\cdots\cdots H_1A=\begin{bmatrix}U\\ 0\end{bmatrix}, \tag{43}$$

其中的 U 是一个 $r\times n$ 阶上梯形阵. 由于直交阵的乘积仍为直交阵，记 $Q^T=H_r\cdots\cdots H_1$，则 Q 也是直交阵，于是矩阵 A 便可以表示成

$$A=Q\begin{bmatrix}U\\ 0\end{bmatrix}.$$

由于右端 U 下的 $m-r$ 行全是零元素，所以在运算时 Q 的后 $m-r$ 列实际上不起作用、记 Q 的前 r 列为 Q_r，则上式可表示成

$$A=Q_rU, \tag{44}$$

其中，Q_r 和 U 的秩都是 r. 这就是说，我们用 Householder 变换完成了矩阵 A 的满秩分解.

如果矩阵 A 为列满秩，即，$\mathrm{rank}(A)=n$，则上梯形阵 U

这时就变成一个 $n\times n$ 阶上三角阵 R，而 (44) 变成

$$A=Q_nR. \tag{45}$$

在利用 H 变换消元时，第一步作用于 $m\times n$ 阶矩阵 A，第二步作用于 $(m-1)\times(n-1)$ 阶矩阵 $\hat{A}^{(1)}$，等等，被作用的矩阵，阶数越来越小，这就是说，消元的计算量逐步下降.

关于 H 阵的存放：各次消元所用的 H 阵是分别由 m 维向量 $\boldsymbol{u}_1$，$m-1$ 维向量 $\boldsymbol{u}_2$，……，$m-r+1$ 维向量 $\boldsymbol{u}_r$ 所确定的，因此，只要把生成 H 阵的这些向量存放起来就行了. 在消元过程中，存放矩阵 A 下方元素的单元，分批为零所取代，因此，可用它们所占单元来存放诸 $\boldsymbol{u}_i$ 的分量. 这时唯一的困难是主对角上的元素非零，为了便于应用，可把它们存放到另外的单元，让出地方来存放各次向量的第一个分量.

选主元　为了提高算法的数值稳定性，在利用 H 变换进行满秩分解时，选主元的措施也是十分重要的. 选主元的具体方法如下：

在进行第一步以前，先比较 $\|\boldsymbol{a}_i\|$，$i=1$，……，n 的大小，取其中第一个最大者作为 A 的第一个列向量，必要时应对矩阵 A 作适当的列交换. 完成以后再进行第一步消元.

在第二步开始前，比较矩阵 $\hat{A}^{(1)}=[\hat{\boldsymbol{a}}_2, \cdots\cdots, \hat{\boldsymbol{a}}_n]$ 中 $n-1$ 个 $m-1$ 维列向量的大小，选长度最大者作为 $\hat{A}^{(1)}$ 的第一列，需要作列交换时应对矩阵 $A^{(1)}$ 进行. 完成以后再进行第二步. 如此继续下去. 这样就可以保证最后所得上梯形矩阵 U 中的各主对角元，其绝对值不增，并且在每一次消元后，相应主对角元都具有最大的绝对值.

公式 (44) 和 (45) 便是矩阵 A 的 QU 分解和 QR 分解，它们分别相应于 A 为亏秩和(列)满秩的情形. 完成这种分解，当 $r\gg1$ 时，所需总运算量的阶为：

$$2r\left[mn-\frac{1}{2}(m+n)+\frac{1}{3}r^2\right].$$

它是 Gauss 消去法完成 LU 分解所需运算量的两倍，但是，为完成广义逆矩阵 A^+ 的计算而进行下一阶段工作时，QU 分解却得到了补偿. 例如，利用 Gauss 消去法并采用矩阵 A 的 LDU 分解计算 A^+ 尚需的运算次数约为

$$r\left[2mn+\frac{1}{2}mr+r^2\right],$$

而用 QU 分解去完成 A^+ 的计算则只需约

$$r\left[2mn-\frac{1}{2}mr-nr+\frac{4}{3}r^2\right]$$

次运算.

3-3 极小最小二乘解的计算和 QR–LL^T 分解

在完成矩阵 A 的 QU 分解后，线性方程组

$$A\boldsymbol{x}=\boldsymbol{b}$$

的极小最小二乘解就可以表示成

$$\boldsymbol{x}=U^T(UU^T)^{-1}Q_r^T\boldsymbol{b}. \tag{46}$$

右端中的 $Q_r^T\boldsymbol{b}$ 是一个 r 维向量，它的计算不需要先形成 H 变换的乘积以得到 Q_r 后再进行，在对矩阵进行分解的过程中，随即依次用 H 变换 H_1, H_2, ……, H_r 左乘向量 $\boldsymbol{b}$，最后取其前 r 个分量构成 r 维向量 $Q_r^T\boldsymbol{b}$. 记其为 $\boldsymbol{b}_r$. 于是，(46) 就变成

$$\boldsymbol{x}=U^T(UU^T)^{-1}\boldsymbol{b}_r. \tag{47}$$

在计算极小最小二乘解 $\boldsymbol{x}$ 时，并不需要把逆矩阵 $(UU^T)^{-1}$ 先计算出来，可按下列方式进行：令

$$(UU^T)^{-1}\boldsymbol{b}_r=\boldsymbol{v},$$

则

$$(UU^T)\boldsymbol{v}=\boldsymbol{b}_r. \tag{48}$$

因为 UU^T 为一个 $r\times r$ 阶对称正定矩阵，所以可用 Cholesky

分解法来计算向量 $\boldsymbol{v}$, 代入 (47) 中就得到了所需要的解 $\boldsymbol{x}$.

若 $r=n$, 即矩阵 A 为列满秩, 这时 $U=R$ 是一个非奇异的上三角阵, 于是, (47) 就化为

$$\begin{aligned}\boldsymbol{x}&=R^T(RR^T)^{-1}\boldsymbol{b}_n=R^T(R^T)^{-1}R^{-1}\boldsymbol{b}_n\\&=R^{-1}\boldsymbol{b}_n,\end{aligned}\tag{49}$$

或

$$R\boldsymbol{x}=\boldsymbol{b}_n.\tag{50}$$

只需利用回代就可以把解 $\boldsymbol{x}$ 计算出来.

QR-LL^T 分解法: 在一般情形, 矩阵 A 的秩 r 常和 n 很接近, 即

$$n-r=q$$

是一个小的自然数. 对于这种情形, 应用以下的方法比较理想.

把矩阵 A 的 QU 分解中的 $r\times n$ 阶上梯形矩阵 U 表示成

$$U=[\tilde{R}\vdots\tilde{U}],\tag{51}$$

这里的 $\tilde{R}$ 是一个非奇异的 $r\times r$ 阶上三角阵, $\tilde{U}$ 是一般的 $r\times(n-r)$ 阶矩阵. 把向量 $\boldsymbol{x}$ 分成 r 维和 $n-r$ 维两个部分, 记之为 $\boldsymbol{x}_1$ 和 $\boldsymbol{x}_2$,

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_1\\\boldsymbol{x}_2\end{bmatrix}.\tag{52}$$

用 $U=[\tilde{R}\tilde{U}]$ 左乘 (47) 的两端, 便得到方程

$$\tilde{R}\boldsymbol{x}_1+\tilde{U}\boldsymbol{x}_2=\boldsymbol{b}_r.\tag{53}$$

这是含 n 个未知量的 r 个线性方程. 从 (53) 解出 $\boldsymbol{x}_1$

$$\boldsymbol{x}_1=\tilde{R}^{-1}(\boldsymbol{b}_r-\tilde{U}\boldsymbol{x}_2),\tag{54}$$

然后再确定 $\boldsymbol{x}_2$ 使

$$\begin{aligned}\|\boldsymbol{x}\|^2&=\|\boldsymbol{x}_1\|^2+\|\boldsymbol{x}_2\|^2\\&=[\tilde{R}^{-1}(\boldsymbol{b}_r-\tilde{U}\boldsymbol{x}_2)]^T[\tilde{R}^{-1}(\boldsymbol{b}_r-\tilde{U}\boldsymbol{x}_2)]+\boldsymbol{x}_2^T\boldsymbol{x}_2=\text{极小}.\end{aligned}$$

令 $\text{grad}(\|\boldsymbol{x}\|^2)=0$, 便得到方程

$$[\tilde{U}^T(\tilde{R}\tilde{R}^T)^{-1}\tilde{U}+I_{n-r}]\boldsymbol{x}_2=\tilde{U}^T(\tilde{R}\tilde{R}^T)^{-1}\boldsymbol{b}_r,$$

其中，I_{n-r} 是$(n-r)\times(n-r)$阶单位阵．把这后一方程改写为

$$[(\tilde{R}^{-1}\tilde{U})^T(\tilde{R}^{-1}\tilde{U})+I_{n-r}]\boldsymbol{x}_2=(\tilde{R}^{-1}\tilde{U})^T\tilde{R}^{-1}\boldsymbol{b}_r. \tag{55}$$

令 $V=\tilde{R}^{-1}\tilde{U}$，方程 (55) 就简化为：

$$(I_{n-r}+V^TV)\boldsymbol{x}_2=V^T\tilde{R}^{-1}\boldsymbol{b}_r. \tag{56}$$

矩阵 V 是 $r\times(n-r)$ 阶，根据方程

$$\tilde{R}V=\tilde{U} \tag{57}$$

依次进行 $n-r$ 次回代就可以把它计算出来．

方程 (56) 的系数矩阵 $I_{n-r}+V^TV$ 是一个 $(n-r)\times(n-r)$ 阶对称正定方阵，而 $n-r$ 又很小，很容易作出它的 Cholesky 分解：

$$I_{n-r}+V^TV=LL^T, \tag{58}$$

L 是一个 $(n-r)\times(n-r)$ 阶下三角阵．

令 $\tilde{R}^{-1}\boldsymbol{b}_r=\boldsymbol{y}$，则

$$\tilde{R}\boldsymbol{y}=\boldsymbol{b}_r,$$

经回代即可算出向量 $\boldsymbol{y}$．然后再解一个小方程

$$LL^T\boldsymbol{x}_2=V^T\boldsymbol{y}, \tag{59}$$

最后便得到了所需要的极小最小二乘解

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{y}-V\boldsymbol{x}_2\\ \boldsymbol{x}_2\end{bmatrix}. \tag{60}$$

生成方程 (59)，总共约需

$$n-r\text{ 次开方，}\quad \frac{1}{2}(n-r)r^2\text{ 次乘法，}$$

运算量不大，所以当 $n-r$ 很小时是十分方便的．

3-4 QU 分解的误差分析和确定伪秩的准则

在前面所讲的算法中涉及到矩阵 A 的秩，这是事先并不知道的，应在算法执行过程中确定一个理想的伪秩．在第 4

章中我们对确定伪秩提出了一个准则，根据有关的相关性指标的变化情况来确定伪秩的大小．本节将介绍 Deuflhard 和 Sautter 提出的一种误差分析方法，并据此阐明前述准则的合理性．

记初始矩阵 A 为 $A^{(0)}$，H_k 为第 k 次的 Householder 变换，则有

$$\begin{aligned} &A=A^{(0)},\ A^{(k)}=H_kA^{(k-1)},\ k=1,\ \cdots\cdots,\ n, \\ &\boldsymbol{b}^{(0)}=\boldsymbol{b},\ \boldsymbol{b}^{(k)}=H_k\boldsymbol{b}^{(k-1)},\ k=1,\ \cdots\cdots,\ n. \end{aligned} \tag{61}$$

这里暂假定 A 为满秩．k 次 H 变换的结果，把矩阵 $A^{(0)}$ 化为以下的形式：

$$A^{(k)}=H_kH_{k-1}\cdots H_1A^{(0)}=\begin{bmatrix} R_k & U_k \\ 0 & W_k \end{bmatrix}, \tag{62}$$

相应的右端向量为：

$$\boldsymbol{b}^{(k)}=[\boldsymbol{c}_k^T\ \boldsymbol{d}_k^T]^T. \tag{63}$$

为了记号简单，我们假定矩阵 A 中各列的次序已符合选主元的要求，不需要右乘以排列阵作列交换．于是，上三角阵 R_k 中的主对角元 $|\rho_{ii}|$，为一不增序列：

$$|\rho_{11}|\geqslant|\rho_{22}|\geqslant\cdots\cdots\geqslant|\rho_{kk}|,$$

同时有
$$|\rho_{ii}|\geqslant|\rho_{ij}|,\ 对\ j>i.$$

为了进行误差分析的需要，今规定一种特殊的矩阵范数如下：

$$\|A\|_\square\equiv\max_i\|\boldsymbol{a}_i\|,\ A=[\boldsymbol{a}_1,\ \cdots\cdots,\ \boldsymbol{a}_n]. \tag{64}$$

根据这个定义，分解式 (62) 中有关矩阵的 $\|\cdot\|_\square$ 如下：

$$\left.\begin{aligned} &\|A\|_\square=|\rho_{11}|, \\ &\|U_k\|_\square\leqslant\|R_k\|_\square=|\rho_{11}|, \\ &\|W_k\|_\square=|\rho_{k+1,k+1}|, \\ &\|W_{k+1}\|_\square\leqslant\|W_k\|_\square, \end{aligned}\right\} k=1,\ \cdots\cdots,\ n-1. \tag{65}$$

而 (63) 中的向量 $\bar{\boldsymbol{d}}_k$ 满足关系 $\|\bar{\boldsymbol{d}}_{k+1}\|\leqslant\|\bar{\boldsymbol{d}}_k\|$, $k=1, \cdots, n-1$.

若矩阵 A 为亏秩, 则对某一个自然数 $p<n$, 有

$$W_p=0.$$

在矩阵分解进行的过程中, 由于舍入误差的影响, 每一次所作 H 变换的结果, 都带有一定的摄动. 首先, 矩阵 A 记入计算机时带有误差, 实际输入的是摄动矩阵

$$\bar{A}^{(0)}=A+\delta A.$$

设第 k 次精确的 H 变换为 $\bar{H}_k$, 它作用于前次所得矩阵 $\bar{A}^{(k-1)}$ 的结果, 相当于作用于某一摄动矩阵 $\bar{A}^{(k-1)}+F_k$, 即

$$\bar{A}^{(k)}=\bar{H}_k(\bar{A}^{(k-1)}+F_k),\ k=1, \cdots, n. \tag{66}$$

右端向量的计算也相似:

$$\begin{aligned}&\bar{\boldsymbol{b}}^{(0)}=\boldsymbol{b}+\delta\boldsymbol{b},\\&\bar{\boldsymbol{b}}^{(k)}=\bar{H}_k(\bar{\boldsymbol{b}}^{(k-1)}-\boldsymbol{f}_k),\ k=1, \cdots, n.\end{aligned} \tag{67}$$

矩阵 $\bar{A}^{(k)}$ 的形式为:

$$\bar{A}^{(k)}=\begin{bmatrix}\bar{R}_k & \bar{U}_k\\ 0 & \bar{W}_k\end{bmatrix},\quad \bar{\boldsymbol{b}}^{(k)}=\begin{bmatrix}\bar{\boldsymbol{c}}_k\\ \bar{\boldsymbol{d}}_k\end{bmatrix}. \tag{68}$$

这时, 范数关系 (65) 近似地化为:

$$\begin{aligned}&\|\bar{A}^{(0)}\|_{\square}\approx|\bar{\rho}_{11}|,\\&\|\bar{U}_k\|_{\square}\leqslant\|\bar{R}_k\|_{\square}\approx|\bar{\rho}_{11}|,\\&\|\bar{W}_k\|_{\square}\approx|\bar{\rho}_{k+1,k+1}|,\\&\|\bar{W}_{k+1}\|_{\square}\leqslant\|\bar{W}_k\|_{\square},\end{aligned} \tag{69}$$

以及 $\|\bar{\boldsymbol{d}}_{k+1}\|\leqslant\|\bar{\boldsymbol{d}}_k\|$, $k=1, \cdots, n-1$.

上述关系准确到计算机精度的一阶. 根据 Wilkinson 对 H 变换所作的误差分析可得,

$$\|F_k\|_{\square}\leqslant\alpha\eta|\bar{\rho}_{k,k}|,\ \|\boldsymbol{f}_k\|\leqslant\alpha\eta\|\bar{\boldsymbol{d}}_k\|,\ k=1, \cdots, n, \tag{70}$$

其中的 α 是一个常数, 当使用双精度内积累加时, 其值为

12.33，η 表示计算机精度.

对 $k=1, 2, \cdots, p$，利用归纳法便得到下列关系：

$$\begin{aligned}
&\bar{A}^{(p)}=\bar{Q}_p^T[A+\delta A+\bar{F}_p],\ \bar{Q}_p=\bar{H}_1\cdots\bar{H}_p,\\
&\bar{F}_p=F_1+\bar{H}_1F_2+\cdots+\bar{H}_1\cdots\bar{H}_{p-1}F_p,\\
&\bar{\boldsymbol{b}}^{(p)}=\bar{Q}_p^T(\boldsymbol{b}+\delta\boldsymbol{b}+\bar{\boldsymbol{f}}_p),\\
&\bar{\boldsymbol{f}}_p=\boldsymbol{f}_1+\bar{H}_1\boldsymbol{f}_2+\cdots+\bar{H}_1\cdots\bar{H}_{p-1}\boldsymbol{f}_p.
\end{aligned} \tag{71}$$

因此，p 次 H 变换所产生舍入误差的影响，相当于初始摄动

$$A+\delta A+\bar{F}_p,\ \boldsymbol{b}+\delta\boldsymbol{b}+\bar{\boldsymbol{f}}_p \tag{72}$$

所产生的影响. 根据关系 (69)～(71)，便得到摄动(72)的估计：

$$\begin{aligned}
&\|\bar{F}_p\|_E\leqslant\alpha\eta(|\bar{\rho}_{11}|+\cdots+|\bar{\rho}_{pp}|)\leqslant p\alpha\eta\|\bar{A}^{(0)}\|_E,\\
&\|\bar{\boldsymbol{f}}_p\|\leqslant\alpha\eta(\|\bar{\boldsymbol{b}}^{(0)}\|+\|\bar{\boldsymbol{d}}_1\|+\cdots+\|\bar{\boldsymbol{d}}_{p-1}\|)\leqslant p\alpha\eta\|\bar{\boldsymbol{b}}^{(0)}\|.
\end{aligned} \tag{73}$$

于是，按照向后误差分析的观点来看，QU 分解法是数值稳定的.

现在我们来进一步考虑舍入误差在中间计算过程中的影响，并据以建立确定矩阵伪秩的准则. 我们假定 p 次 H 变换作用于 $A+\delta A$ 的精确结果为

$$\hat{A}^{(p)}=\bar{Q}_p^T(A+\delta A)=\begin{bmatrix}\hat{R}_p & \hat{U}_p\\ \delta\widehat{W}_p & \widehat{W}_p\end{bmatrix},$$

$$\bar{Q}_p=\bar{H}_1\cdots\bar{H}_p.$$

以后的 H 变换都只作用于 $\bar{W}_p$ 或 $\widehat{W}_p$. 如果想要计算能继续向下进行，一个合理的准则应是相对误差

$$\frac{\|\bar{W}_p-\widehat{W}_p\|_E}{\|\bar{W}_p\|_E}<1 \tag{74}$$

即

$$\|\bar{W}_p-\widehat{W}_p\|_E<\|\bar{W}_p\|_E.$$

如果 (74) 已不复成立，就说明舍入误差的影响已使 $\bar{W}_p$ 的精度完全丧失，即，这时 $\bar{W}_p$ 中的元素已为舍入误差所控制，它

就已失掉实际意义，从而便没有向下进行的必要了．根据 $\overline{A}^{(p)}$ 和 $\hat{A}^{(p)}$ 的定义，不难推出

$$\|\overline{W}_p-\widehat{W}_p\|_F\leqslant\|\overline{A}_p-\hat{A}_p\|_F=\|F_p\|_F. \tag{75}$$

根据 (73)，可以把条件 (74) 化为一个适宜于计算的条件：

$$\frac{\|\overline{W}_p-\widehat{W}_p\|_F}{\|\overline{W}_p\|_F}\leqslant\frac{\alpha\eta(|\rho_{11}|+\cdots+|\rho_{pp}|)}{|\rho_{p+1,p+1}|}<1. \tag{76}$$

这里的 ρ_{ii} 是从实际计算得到的，一般说来，中间分子括号中的第一项 $|\rho_{11}|$ 占主导地位，从而 (76) 可以表示成

$$\frac{\|\overline{W}_p-\widehat{W}_p\|_F}{\|\overline{W}_p\|_F}=\frac{\varepsilon|\rho_{11}|}{|\rho_{p+1,p+1}|}<1, \tag{77}$$

这里的 ε 是区间 $[\alpha\eta, \alpha p\eta]$ 中的某一个数．这就表明，因子 $|\rho_{11}|/|\rho_{p+1,p+1}|$ 反映舍入误差积累的程度．据此便可以给出确定矩阵伪秩的准则：

给定一适当正数 ε，若自然数 r 满足不等式

$$|\rho_{r+1,r+1}|\leqslant\varepsilon|\rho_{11}|<|\rho_{rr}|, \tag{78}$$

则可确定矩阵的伪秩为 r(ε–秩)．

关于确定伪秩的问题，有几点应加以说明：

(i) 在叙述一个算法的基本思想或步骤时，我们常说相应的矩阵 A 为满秩或亏秩，这仅仅是为了方便，实际上是并不确切的．矩阵 A 即使在理论上为满秩，但由于病态性质比较严重，如果作为满秩处理，所得计算结果将不堪设想．而确定一个适当的伪秩，常能改善问题的条件，得到比较好的结果．

(ii) 在实际计算时，常会发生这样的情形：$|\rho_{k,k}|$ 还比较大，而 $|\rho_{k+1,k+1}|$ 却突然变得很小，这是问题条件较好的反映，这时，确定秩为 k 可以得到好的计算结果．如果序列 $\{|\rho_{ii}|\}$ 缓慢下降，这反映问题的病态性质比较严重．在必要

时应考虑问题的物理背景，重建数学模型.

(iii) 控制量 ε 的选择，对确定伪秩十分重要，ε 取得小，所定伪秩就大，反之，ε 取得大，所定伪秩就小，一般可取略低于计算机精度的量为 ε.

§4 直交化方法

4-1 改进的 Gram-Schmidt 直交化方法

在前一节中我们着重介绍了用 H 变换对矩阵 A 作 QU 或 QR 分解. 实际上，如果对矩阵 $A=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n]$ 的 n 个列向量用直交化方法产生一个标准直交系 $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n$，使每一个 $\boldsymbol{q}_i$ 都是 A 的前 i 个列向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_i$ 的线性组合，并和前 $i-1$ 个列向量都直交，即得到 $m\times n$ 阶列直交矩阵

$$Q=[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n]. \tag{79}$$

在这里我们暂假定 $\mathrm{rank}(A)=n$. 用 Q^T 左乘 A，就得到

$$Q^TA=\begin{bmatrix} R \\ 0 \end{bmatrix},$$

其中的 R 是一个 $n\times n$ 阶上三角阵，下方的 0 是 $(m-n)\times n$ 阶零矩阵. 从而便得到了矩阵 A 的 QR 分解

$$A=QR. \tag{80}$$

因此，从原则上讲，用向量系的直交化方法也可以求得矩阵的 QR 分解. 但是，前面曾经指出，古典的 Gram-Schmidt 直交化方法的数值稳定性不好，本节将介绍它的一种变形，即所谓改进的 Gram-Schmidt 直交化方法，此法具有和 H 变换同等的数值稳定性. 步骤如下：

记

$$A=A^{(1)}=[\boldsymbol{a}_1^{(1)}, \cdots, \boldsymbol{a}_n^{(1)}],\ \boldsymbol{a}_i^{(1)}=\boldsymbol{a}_i,\ i=1, \cdots, n.$$

第一步：令 $u_{11}=\|\boldsymbol{a}_1^{(1)}\|$,

$$\boldsymbol{q}_1=\boldsymbol{a}_1^{(1)}/u_{11}.$$

依次令
$$\boldsymbol{a}_i^{(2)}=\boldsymbol{a}_i^{(1)}-u_{2i}\boldsymbol{q}_1,\ i=2,\ 3,\ \cdots,\ n,$$
并确定系数 u_{2i} 使

$$\boldsymbol{a}_i^{(2)}\perp\boldsymbol{q}_1,\ i=2,\ \cdots,\ n,$$

即
$$u_{2i}=\boldsymbol{q}_1^T\boldsymbol{a}_i^{(1)},\ i=2,\ \cdots,\ n.$$

令 $u_{22}=\|\boldsymbol{a}_2^{(2)}\|$, $\boldsymbol{q}_2=\boldsymbol{a}_2^{(2)}/u_{22}$, 于是 $A^{(1)}$ 就化为

$$A^{(2)}=[\boldsymbol{q}_1,\ \boldsymbol{q}_2,\ \boldsymbol{a}_3^{(2)},\ \cdots,\ \boldsymbol{a}_n^{(2)}].$$

第二步：仿前一步的办法，令

$$\boldsymbol{a}_i^{(3)}=\boldsymbol{a}_i^{(2)}-u_{3i}\boldsymbol{q}_2,\ i=3,\ 4,\ \cdots,\ n,$$

其中的 u_{3i} 应为

$$u_{3i}=\boldsymbol{q}_2^T\boldsymbol{a}_i^{(2)},\ i=3,\ \cdots,\ n.$$

并令 $u_{33}=\|\boldsymbol{a}_3^{(3)}\|$, $\boldsymbol{q}_3=\boldsymbol{a}_3^{(3)}/u_{33}$. 于是就完成了第二步，这时 $A^{(2)}$ 化为

$$A^{(3)}=[\boldsymbol{q}_1,\ \boldsymbol{q}_2,\ \boldsymbol{q}_3,\ \boldsymbol{a}_4^{(3)},\ \cdots,\ \boldsymbol{a}_n^{(3)}].$$

在完成第 $s-1$ 步后，矩阵 $A^{(s-1)}$ 就化为：

$$A^{(s)}=[\boldsymbol{q}_1,\ \cdots,\ \boldsymbol{q}_s,\ \boldsymbol{a}_{s+1}^{(s)},\ \cdots,\boldsymbol{a}_n^{(s)}],$$

其中

$$u_{ss}=\|\boldsymbol{a}_s^{(s)}\|,\ \boldsymbol{q}_s=\boldsymbol{a}_s^{(s)}/u_{ss},$$
$$u_{si}=\boldsymbol{q}_{s-1}^T\boldsymbol{a}_i^{(s-1)},\ i=s+1,\ \cdots,\ n,$$

于是，经 $n-1$ 步后就得了所需要的直交系

$$A^{(n)}=[\boldsymbol{q}_1,\ \cdots,\ \boldsymbol{q}_n].$$

4-2 选主元与伪秩的确定

为了加强算法的数值稳定性，在直交化过程中也常采用选主元的方法．其法如下：

在对 $A^{(1)}=A=[\boldsymbol{a}_1^{(1)},\ \cdots,\ \boldsymbol{a}_n^{(1)}]$ 进行第一步直交化开始，先选取向量系 $\{\boldsymbol{a}_i^{(1)}\}$ 中的最长者作为第一个向量．今仍记 $\boldsymbol{i}$

为 $\boldsymbol{a}_1^{(1)}$, 并令

$$u_{11}=\|\boldsymbol{a}_1^{(1)}\|,\ \boldsymbol{q}_1=\boldsymbol{a}_1^{(1)}/u_{11}.$$

在第 $s-1$ 步直交化完成后得到了

$$A^{(s)}=[\boldsymbol{q}_1,\ \cdots,\ \boldsymbol{q}_{s-1},\ \boldsymbol{a}_s^{(s)},\ \cdots,\ \boldsymbol{a}_n^{(s)}].$$

在向量系 $\boldsymbol{a}_s^{(s)}$, $\cdots$, $\boldsymbol{a}_n^{(s)}$ 中取其最长者作为 $A^{(s)}$ 的第 s 列, 并仍记之为 $\boldsymbol{a}_s^{(s)}$, 令

$$u_{ss}=\|\boldsymbol{a}_s^{(s)}\|,\ \boldsymbol{q}_s=\boldsymbol{a}_s^{(s)}/u_{ss},$$

$$u_{si}=\boldsymbol{q}_{s-1}^T\boldsymbol{a}_i^{(s-1)},\ i=s+1,\ \cdots,\ n.$$

这样继续下去, 当 A 为满秩时, 经 $n-1$ 步就完成了直交化工作.

经过选主元后, 序列 $\{|u_{ii}|\}$ 为不增. 若矩阵 A 的列向量已规范化, 则 $|u_{11}|=1$. 这时, $|u_{kk}|$ 就是 $\boldsymbol{a}_k$ 关于 $\boldsymbol{a}_1,\cdots,\boldsymbol{a}_{k-1}$ (指列交换后的情形)的相对相关性指标. 在应用直交化方法时, 也要确定伪秩, 其原则和前节一样.

4-3 广义逆矩阵和极小最小二乘解的计算

设矩阵 A 的伪秩已确定为 r, 则用直交化方法便得到了分解

$$A=Q_rU, \tag{81}$$

其中, Q_r 是一个 $m\times r$ 阶列直交矩阵, U 是一个 $r\times n$ 阶矩阵, 二者均为满秩. 这时, 若不考虑算法中的列交换,

$$A^+=U^T(UU^T)^{-1}Q_r^T, \tag{82}$$

而极小最小二乘解为

$$\boldsymbol{x}=U^T(UU^T)^{-1}Q_r^T\boldsymbol{b}. \tag{83}$$

关于 $Q_r^T\boldsymbol{b}$ 的计算可按以下的方式进行.

$Q_r^T\boldsymbol{b}$ 是一个 r 维向量, 它的 r 个分量为

$$\boldsymbol{q}_1^T\boldsymbol{b},\ \boldsymbol{q}_2^T\boldsymbol{b},\ \cdots,\ \boldsymbol{q}_r^T\boldsymbol{b}.$$

若用 c_i 表示 $\boldsymbol{q}_i^T\boldsymbol{b}$, $i=1, 2, \cdots, r$, 其值可依次计算如下：令

$$\begin{aligned}\boldsymbol{b}&=\boldsymbol{b}_1,\ c_i=\boldsymbol{q}_i^T\boldsymbol{b}_i,\\ \boldsymbol{b}_{i+1}&=\boldsymbol{b}_i-c_i\boldsymbol{q}_i,\end{aligned}\tag{84}$$

于是 c_i 和 $\boldsymbol{b}_i$ 便都交替计算出来. 此法效果较好，参阅文献[9].

4-4 直交化方法的数值稳定性问题

关于古典的 Gram–Schmidt 直交化方法和它的变形，改进的直交化方法的数值稳定性问题，Rice 在 1966 年发表了一个数值实验报告，并作了一个简单的误差分析，说明古典的直交化方法的数值稳定性不好，往往需要进行第二次直交化. 而改进的直交化方法则具有较好的数值稳定性、Björk 在 1967 年发表了所作的详细的误差分析. 对于用改进的直交化方法来计算超定线性方程组的极小最小二乘解，若使用双精度内积累加，改进的直交化方法得到的结果是某一摄动方程

$$(A+\delta A)\boldsymbol{x}=\boldsymbol{b}+\delta\boldsymbol{b}\tag{85}$$

的准确解. 摄动量 δA 和 $\delta\boldsymbol{b}$ 满足不等式

$$\begin{aligned}\|\delta A\|_E&\leqslant 2n^{\frac{3}{2}}\|A\|_E\eta,\\ \|\delta\boldsymbol{b}\|&\leqslant 2n^{\frac{3}{2}}\|\boldsymbol{b}\|\eta,\end{aligned}\tag{86}$$

其中 $\|\cdot\|_E$ 表示

$$\|A\|_E=(\sum a_{ij}^2)^{1/2},\tag{87}$$

称为矩阵 A 的欧氏范数或 Frobenius 范数.

用直交化方法完成矩阵的 QU 分解的运算量约为

$$2rm(n-\frac{1}{2}r),$$

这里的运算是指一次乘法和一次加法. 它比用 H 变换要略大一些. 所需存贮量也是这样，这是因为用 H 变换进行分解时，确定 H 变换的向量维数逐次下降，不但减少了运算量，同

时也使存贮量相应减小.

§5 计算广义逆矩阵和极小最小二乘解的递推方法

5-1 Greville 递推方法——矩阵满秩的情形

在本节中我们仍记

$$A_k=[\boldsymbol{a}_1, \cdots, \boldsymbol{a}_k],\ k=1, \cdots, n. \tag{88}$$

所谓递推方法, 就是在 A_k^+ 算出以后, 根据一种递推关系直接把 A_{k+1}^+ 算出来. 这种递推方法, 实际上是 Gram-Schmidt 直交化方法的一种实现形式. 为了说明它们之间的关系, 我们的讨论就从矩阵的 QR 分解着手. 首先我们考虑矩阵 A_n 为满秩的情形. 对 A_n 作 QR 分解

$$A_n=Q_nR_n, \tag{89}$$

其中的 Q_n 是一个 $m\times n$ 阶列直交矩阵, R_n 是一个非奇异的 $n\times n$ 阶上三角阵. 令 $Q_n=[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_n]$. 由此可知

$$A_k=Q_kR_k, \tag{90}$$

其中 $Q_k=[\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k]$, 而 R_k 是 R_n 中位于左上角的那个 $k\times k$ 阶上三角子矩阵. 根据 A_k 的分解式 (90) 可得

$$A_k^+=R_k^{-1}Q_k^T. \tag{91}$$

设 $k\leqslant n-1$, 因为

$$A_{k+1}=[A_k\ \vdots\ \boldsymbol{a}_{k+1}],\ Q_{k+1}=[Q_k\ \vdots\ \boldsymbol{q}_{k+1}], \tag{92}$$

且

$$R_{k+1}=\begin{bmatrix} R_k & \boldsymbol{r}_{k+1} \\ 0 & \alpha_{k+1} \end{bmatrix}, \tag{93}$$

其中的 $\boldsymbol{r}_{k+1}$ 表示一个 k 维列向量, α_{k+1} 是一个数, 不难验证

$$R_{k+1}^{-1}=\begin{bmatrix} R_k^{-1} & -R_k^{-1}\boldsymbol{r}_{k+1}/\alpha_{k+1} \\ 0 & 1/\alpha_{k+1} \end{bmatrix}, \tag{94}$$

从而有

$$\boldsymbol{r}_{k+1}=Q_k^T\boldsymbol{a}_{k+1},\ \alpha_{k+1}=\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1}. \tag{95}$$

于是就可把 R_{k+1}^{-1} 具体表示成

$$R_{k+1}^{-1}=\begin{bmatrix} R_k^{-1} & -R_k^{-1}Q_k^T\boldsymbol{a}_{k+1}/\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1} \\ 0 & 1/\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1} \end{bmatrix}. \tag{96}$$

据此，我们就得到表达式

$$A_{k+1}^+=R_{k+1}^{-1}Q_{k+1}^T=\begin{bmatrix} R_k^{-1}Q_k^T-R_k^{-1}Q_k^T\boldsymbol{a}_{k+1}\boldsymbol{q}_{k+1}^T\dfrac{1}{\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1}} \\ \dfrac{1}{\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1}}\boldsymbol{q}_{k+1}^T \end{bmatrix}. \tag{97}$$

联系到前节中的直交化方法，向量 $\boldsymbol{q}_{k+1}$ 是由向量 $\boldsymbol{a}_{k+1}$ 在 $R^\perp(A_k)$ 上的直交投影 $\mathbf{c}_{k+1}$ 而生成的，即

$$\mathbf{c}_{k+1}=(I_m-A_kA_k^+)\boldsymbol{a}_{k+1} \tag{98}$$

而

$$\boldsymbol{q}_{k+1}=\frac{1}{\|\mathbf{c}_{k+1}\|}\mathbf{c}_{k+1}. \tag{99}$$

又，因为 $\mathbf{c}_{k+1}\neq\mathbf{0}$，它的广义逆为

$$\mathbf{c}_{k+1}^+=(\mathbf{c}_{k+1}^T\mathbf{c}_{k+1})^{-1}\mathbf{c}_{k+1}^T=\frac{1}{\|\mathbf{c}_{k+1}\|}\boldsymbol{q}_{k+1}^T, \tag{100}$$

但

$$\|\mathbf{c}_{k+1}\|^2=\mathbf{c}_{k+1}^T\boldsymbol{a}_{k+1}=\boldsymbol{a}_{k+1}^T\boldsymbol{q}_{k+1}\|\mathbf{c}_{k+1}\| \tag{101}$$

所以

$$\frac{1}{\boldsymbol{q}_{k+1}^T\boldsymbol{a}_{k+1}}\boldsymbol{q}_{k+1}^T=\mathbf{c}_{k+1}^+.$$

于是，表达式 (97) 便可以表示成

$$A_{k+1}^+=\begin{bmatrix} A_k^+-A_k^+\boldsymbol{a}_{k+1}\mathbf{c}_{k+1}^+ \\ \mathbf{c}_{k+1}^+ \end{bmatrix}. \tag{102}$$

公式 (102) 便是计算广义逆矩阵的按列递推关系。

注意到 $\mathbf{c}_1, \cdots, \mathbf{c}_k$ 都属于 $R(A_k)$，故必有

$$A_kA_k^+\mathbf{c}_i=\mathbf{c}_i,\ i=1,\ \cdots,\ k.$$

这就是说，向量 $\mathbf{c}_1, \cdots, \mathbf{c}_k$ 正是秩为 k 的实对称矩阵 $A_kA_k^+$ 的

与特征值 1 相应的特征向量系，并且是一个直交系，从而便得到 $A_kA_k^+$ 的谱分解公式

$$A_kA_k^+=\boldsymbol{c}_1\boldsymbol{c}_1^++\cdots+\boldsymbol{c}_k\boldsymbol{c}_k^+. \tag{103}$$

联系到 (98)，公式 (103) 可以用来计算 $\boldsymbol{c}_{k+1}$.

5-2　Greville 递推方法—矩阵亏秩的情形

如果

$$\operatorname{rank}(A_n)=r<n,\ \operatorname{rank}(A_r)=r, \tag{104}$$

则矩阵 A_n 的分解式为

$$A_n=Q_r\widetilde{U}_r,$$

其中的 $\widetilde{U}_r$ 是一个 $r\times n$ 阶上梯形阵，它的形式为:

$$\widetilde{U}_r=[R_r\ U_r], \tag{105}$$

这里的 R_r 是一个 $r\times r$ 阶上三角阵，U_r 是一个 $r\times(n-r)$ 阶矩阵. 在算出 A_r^+ 以后，便可以利用 QR-LL^T 分解法直接把 A_n^+ 算出来. 计算 A_n^+ 的公式如下: 因为

$$U_r=Q_r^T[\boldsymbol{a}_{r+1},\ \cdots,\ \boldsymbol{a}_n],$$

令

$$\begin{aligned}V_r&=R_r^{-1}U_r=A_r^+[\boldsymbol{a}_{r+1},\ \cdots,\ \boldsymbol{a}_n],\\ I_{n-r}&+V_r^TV_r=L_rL_r^T,\end{aligned} \tag{106}$$

则可以把 A_n^+ 表示成

$$A_n^+=\begin{bmatrix}R_r^{-1}Q_r^T-R_r^{-1}U_r(L_rL_r^T)^{-1}V_r^TR_r^{-1}Q_r\\ (L_rL_r^T)^{-1}V_r^TR_r^{-1}Q_r^T\end{bmatrix}. \tag{107}$$

如果把 $[\boldsymbol{a}_{r+1},\ \cdots,\ \boldsymbol{a}_n]$ 记为 $A_{r,n}$，由于 $V_r=R_r^{-1}Q_r^TA_{r,n}$，所以有

$$V_r=A_r^+A_{r,n}. \tag{108}$$

从而可以把 A_n^+ 表示成

$$A_n^+=\begin{bmatrix}B_n\\ C_n\end{bmatrix}, \tag{109}$$

其中

$$B_n = A_r^+ - V_r C_n,$$
$$C_n = (I_{n-r} + V_r^T V_r)^{-1} V_r^T A_r^+. \tag{110}$$

5-3 选主元与伪秩的确定

在用递推方法计算广义逆矩阵A_r^-时，为了加强算法的数值稳定性，也需要应用选主元的方法. 和改进的Gram-Schmidt直交化方法中选主元的情形一样，首先在$\boldsymbol{a}_1^{(1)}, \cdots, \boldsymbol{a}_n^{(1)}$中取长度最大的作为第一个列向量$\boldsymbol{a}_1^{(1)}$，令

$$\mathbf{c}_1 = \boldsymbol{a}_1^{(1)}, \tag{111}$$

计算 $\boldsymbol{a}_i^{(2)} = (I - \mathbf{c}_1\mathbf{c}_1^+)\boldsymbol{a}_i^{(1)},\ i=2, \cdots, n,$

选其中的长度最大者作为$\boldsymbol{a}_2^{(2)}$，并令

$$\mathbf{c}_2 = \boldsymbol{a}_2^{(2)}.$$

设$k \leqslant r-1$，并已算出了$\mathbf{c}_k$，计算

$$\boldsymbol{a}_i^{(k+1)} = (I - \mathbf{c}_k\mathbf{c}_k^+)\boldsymbol{a}_i^{(k)},\ i=k+1, \cdots, n. \tag{112}$$

取其中长度最大者作为$\boldsymbol{a}_{k+1}^{(k+1)}$，并令

$$\mathbf{c}_{k+1} = \boldsymbol{a}_{k+1}^{(k+1)}. \tag{113}$$

在选主元的过程中，必要时要对矩阵作相应的列交换. 因为$\operatorname{rank}(A_n)=r$，在算出$\mathbf{c}_r$后选主元的过程就算完成. 由于矩阵的秩事先并不知道，要在选主元的过程中来确定. 容易看出，向量$\mathbf{c}_k$实际上就是向量$\boldsymbol{a}_k$在$R^{\perp}(A_{k-1})$上的直交投影(矩阵中的列向量顺序由于选主元可能发生变化)，所以其长度$\|\mathbf{c}_k\|$便是向量$\boldsymbol{a}_k$关于$\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}$的相对相关性指标，对给定的控制量$\varepsilon$，若不等式

$$\|\mathbf{c}_{k+1}\| \leqslant \varepsilon \|\mathbf{c}_1\| < \|\mathbf{c}_k\| \tag{114}$$

成立，就确定矩阵的伪秩$r=k$.

5-4 按行递推和极小最小二乘解的计算

如果线性方程组非超定，即$m \leqslant n$，这时以按行递推为宜. 把系数矩阵中的行改成列，并记之为

$$\hat{A}_m = A_n^T = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m], \tag{115}$$

右端的向量 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_m$ 是矩阵 A_n 的 m 行转置而成的. 在这种情形, 我们可以应用前面讲过的按列递推方法来计算 $\hat{A}_m^+$, 根据转置关系

$$A_n^+ = (\hat{A}_m^+)^T, \tag{116}$$

最后就得到了所需要的广义逆矩阵 A_n^+. 我们仍分满秩与亏秩两种情形来考虑.

满秩情形: 据 (102),

$$\hat{A}_{k+1}^+ = \begin{bmatrix} \hat{A}_k^+ - \hat{A}_k^+ \boldsymbol{a}_{k+1} \mathbf{c}_{k+1}^+ \\ \mathbf{c}_{k+1}^+ \end{bmatrix}, \tag{117}$$

$$\hat{A}_1^+ = \boldsymbol{a}_1^+, \ \mathbf{c}_1 = \boldsymbol{a}_1.$$

转置得

$$(\hat{A}_{k+1}^+)^T = [(\hat{A}_k^+)^T - \mathbf{c}_{k+1}^{+T} \boldsymbol{a}_{k+1}^T (\hat{A}_k^+)^T \vdots \mathbf{c}_{k+1}^{+T}]. \tag{118}$$

记方程的右端向量为

$$\boldsymbol{b} = [\beta_1, \cdots, \beta_m]^T,$$

并记其前 k 个分量组成的向量为

$$\boldsymbol{b}_k = [\beta_1, \cdots, \beta_k]^T. \tag{119}$$

记前 i 个方程的极小最小二乘解为 $\boldsymbol{x}_i$,

$$\boldsymbol{x}_i = (\hat{A}_i^T)^+ \boldsymbol{b}_i, \ i = 1, 2, \cdots, m,$$

则据 (118) 得

$$\begin{aligned} \boldsymbol{x}_{k+1} &= (\hat{A}_{k+1}^T)^+ \boldsymbol{b}_{k+1} \\ &= [(\hat{A}_k^T)^+ - \mathbf{c}_{k+1}^{+T} \boldsymbol{a}_{k+1}^T (\hat{A}_k^T)^+ \vdots \mathbf{c}_{k+1}^{+T}] \begin{bmatrix} \boldsymbol{b}_k \\ \beta_{k+1} \end{bmatrix} \\ &= \boldsymbol{x}_k + r_{k+1}(\boldsymbol{x}_k) \mathbf{c}_{k+1}^{+T}, \end{aligned} \tag{120}$$

其中 $r_{k+1}(\boldsymbol{x}_k)$ 表示第 $k+1$ 个方程当 $\boldsymbol{x} = \boldsymbol{x}_k$ 时的残量

$$r_{k+1}(\boldsymbol{x}_k) = \beta_{k+1} - \boldsymbol{a}_{k+1}^T \boldsymbol{x}_k, \tag{121}$$

为简单起见, 记其为 d_{k+1}. 于是便得到了计算 $\boldsymbol{x}_k$ 的递推关系. 当矩阵 $\hat{A}_m$ 的秩等于 m 时, $\boldsymbol{x}_m$ 便是相应线性方程组的极

小最小二乘解. 这时, 由于 $m \leqslant n$, 所以方程对任意右端向量 $\boldsymbol{b}$ 都相容, 所以, x_m 在这种情况下也是所给相容线性方程组的极小范数解.

亏秩的情形: 应当指出, 这里所用"亏秩"一词, 并不全面, 因为, 即使系数矩阵在理论上为满秩, 由于问题的病态性质, 也常常需要在算法的执行过程中, 确定一个适当的伪秩, 以提高算法的效果. 为了提高算法的数值稳定性, 这里也采用前述选主元的方法, 并在选主元的过程中把矩阵的伪秩确定下来, 我们假定

$$\boldsymbol{c}_1,\ \boldsymbol{c}_2,\ \cdots,\ \boldsymbol{c}_k$$

已经算出, 现在根据改进的 Gram-Schmidt 直交化方法, 来确定 $\boldsymbol{c}_{k+1}$, 对 $j=k+1,\ \cdots,\ m$, 依次计算

$$\boldsymbol{c}_j^{(k+1)}=\boldsymbol{c}_j^{(k)}-\frac{\boldsymbol{c}_k^T\boldsymbol{c}_j^{(k)}}{\boldsymbol{c}_k^T\boldsymbol{a}_k}\boldsymbol{c}_k, \tag{122}$$

选取其中的长度最大者作为 $\boldsymbol{c}_{k+1}$,

$$\boldsymbol{c}_{k+1}=\boldsymbol{c}_{k+1}^{(k+1)}.$$

这时需要适当的列交换, 对原线性方程组来说, 是相应的行交换, 只需注意右端向量的分量也应作相应的交换, 而 x 分量的次序不变. 为了根据递推关系 (120) 计算 x_{k+1}, 我们可按递推关系

$$d_{k+1}=\beta_{k+1}-\sum_{i=1}^{k}(\boldsymbol{c}_i^T\boldsymbol{a}_{k+1}/\boldsymbol{c}_i^T\boldsymbol{a}_i)d_i, \tag{123}$$

来计算残量 d_{k+1}. 对给定的控制量 ε, 当

$$\|\boldsymbol{c}_{r+1}\|\leqslant\varepsilon$$

时, 就确定矩阵的伪秩为 r, 计算出 x_r 后就完成了第一阶段的计算. 然后, 根据公式 (109) 有

$$\hat{A}_m^+=\begin{bmatrix}\hat{B}_m\\ \hat{C}_m\end{bmatrix}, \tag{124}$$

其中

$$\hat{C}_m=(I_{m-r}+V_r^TV_r)^{-1}V_r^T\hat{A}_r^+;$$
$$\hat{B}_m=\hat{A}_r^+-V_r\hat{C}_m. \tag{125}$$

于是，方程组 $\hat{A}_m^T\boldsymbol{x}=\boldsymbol{b}$ 的极小最小二乘解 $\boldsymbol{x}_m$ 就可以根据公式

$$\boldsymbol{x}_m=\hat{A}_m^{+T}\boldsymbol{b}=[\hat{A}_r^{+T}-\hat{C}_m^TV_r^T \vdots \hat{C}_m^T]\begin{bmatrix}\boldsymbol{b}_r\\ \boldsymbol{b}_{r,m}\end{bmatrix}, \tag{126}$$

其中 $\boldsymbol{b}_{r,m}=[\beta_{r+1}, \cdots, \beta_m]^T$. 此公式还可以表示成

$$\begin{aligned}\boldsymbol{x}_m&=\hat{A}_r^{+T}\boldsymbol{b}_r-\hat{C}_m^TV_r^T\boldsymbol{b}_r+\hat{C}_m^T\boldsymbol{b}_{r,m}\\ &=\boldsymbol{x}_r-\hat{C}_m^TV_r^T\boldsymbol{b}_r+\hat{C}_m^T\boldsymbol{b}_{r,n}.\end{aligned} \tag{127}$$

在参考文献 [12] 中曾对以 Hilbert 矩阵 H_n 为系数矩阵的线性方程组，就 $n=7, 10, 20, 30, 40, 80$ 和100七种情形进行了计算，并对本节的算法和 Cholesky 方法进行了对比。总的情形见下表：

n	Cholesky 法	递推算法	伪秩
7	6 位有效数字	6 位有效数	7
10	无一位有效数	5 位有效数	10
20	无一位有效数	4 位有效数	10
30	无一位有效数	3 位有效数	10
40	无一位有效数	3 位有效数	10
80	无一位有效数	2 位有效数	10
100	无一位有效数	1 位有效数	10

控制量：$\|\boldsymbol{c}_r\|^2\leqslant\varepsilon=10^{-13}$.

就中，$H_{40}\boldsymbol{x}=\boldsymbol{b}$，而 $\boldsymbol{b}$ 的分量为 $b_i=\sum_{j=1}^{40}h_{ij}$，$h_{ij}$ 为 H_{40} 的元素，因此这个方程组的准确解为 $[1, 1, \cdots, 1]^T$. 实际计算结果如下：

0.99990	1.0010	0.99708	1.0012	1.0021
1.0008	0.99945	0.99867	0.99856	0.99891

0.99946	1.0000	1.0004	1.0007	1.0009
1.0009	1.0008	1.0006	1.0004	1.0002
0.99999	0.99979	0.99962	0.99950	0.99962
0.99939	0.99939	0.99944	0.99952	0.99962
0.99974	0.99986	0.99999	1.0001	1.0002
1.0002	1.0003	1.0003	1.0003	1.0002

最后我们谈一下关于控制量 ε 的选择问题. 在上述计算中使用了两个理论上等价的量 $\boldsymbol{c}_k^T\boldsymbol{c}_k$ 和 $\boldsymbol{c}_k^T\boldsymbol{a}_k$, 一般控制在

$$\|\boldsymbol{c}_k\|^2=\boldsymbol{c}_k^T\boldsymbol{c}_k\leqslant 10^{-13}.$$

在使用 $|\boldsymbol{c}_k^T\boldsymbol{a}_k|$ 进行控制时, 由于 $\boldsymbol{c}_k$ 从理论上讲应和 $\boldsymbol{a}_1, \cdots, \boldsymbol{a}_{k-1}$ 都直交, 而实际上是用改进的 Gram-Schmidt 直交化方法计算出来的, 这种直交性并不严格成立. 当 $\|\boldsymbol{c}_k\|$ 变得很小时, 舍入误差的干扰就逐渐明显地表现出来. 基本情况是, 当 $|\boldsymbol{c}_k^T\boldsymbol{a}_k|$ 小于 10^{-12} 时, 以后就开始回升. 而从计算的实际结果来看, 也正是在回升之前所确定的伪秩效果最好. 因此, 我们可以利用这一情况, 对确定伪秩的过程进行自动控制, 而不必先给定控制量 ε.

§6 大型稀疏线性最小二乘问题的计算方法

6-1 引言

在科学技术的许多领域中, 例如, 经济模型, 地震数据分析, 航空测绘以及利用有限元方法进行结构分析中, 常常会遇到大型稀疏线性最小二乘问题. 这一类问题中的系数矩阵 A 不但是稀疏的, 而且还具有一种特殊结构. 因此, 可以利用矩阵的特点来计算问题的极小最小二乘解. 本节所讨论问题中的矩阵 A 来源于大地测量问题, 它具有以下的形式:

$$A = \begin{bmatrix} \text{(block-diagonal hatched blocks)} & 0 & \text{有特殊结构} \\ 0 & \ddots & \end{bmatrix} \tag{128}$$

矩阵 A 为 $m\times n$ 阶，$m\geqslant n$，且 $\mathrm{rank}(A)=n$. 相应的超定方程为

$$A\boldsymbol{x}=\boldsymbol{b} \tag{129}$$

6-2 QR 分解法

矩阵 A 可以表示成

$$A=\begin{bmatrix} A_1 & & & & A_{1,t+1} & \cdots & A_{1,2t-1} \\ & A_2 & & 0 & A_{2,t+1} & \cdots & A_{2,2t-1} \\ & & \ddots & & \vdots & & \vdots \\ & 0 & & \ddots & & & \\ & & & A_t & A_{t,t+1} & \cdots & A_{t,2t-1} \end{bmatrix} \tag{130}$$

在这里 $t=2^k$. 右方的块子矩阵 A_{ij} 中有一些是零块，为考虑方便，对它们暂不加区别. 依次，对每一个 A_i 作 QR 分解

$$Q_iA_i=\begin{bmatrix} R_i \\ 0 \end{bmatrix},\ i=1,\ 2,\ \cdots,\ t, \tag{131}$$

矩阵 Q_i 是一些 $m_i\times m_i$ 阶 H 变换的乘积，R_i 是一些 $n_i\times n_i$ 非奇异上三角阵. 这相当于用 $m\times m$ 阶直交阵 Q^T 来左乘矩阵 A，这里

$$Q=\begin{bmatrix} Q_1 & & & \\ & Q_2 & 0 & \\ & 0 & \ddots & \\ & & & Q_t \end{bmatrix}. \tag{132}$$

应当注意的是，在执行运算时，每一个 Q_i 只和 A_i 所在这 m_i 行中的子块有关，它们是

$$A_i,\ A_{i,t+1}\cdots A_{i,2t-1},$$

而与其它子块无关，所以可节省不少计算量．记

$$Q_i^T[A_i,\ A_{i,t+1},\ \cdots,\ A_{i,2t-1}]$$
$$=\begin{bmatrix} R_i & A_{i,t+1}^{(0)}\cdots A_{i,2t-1}^{(0)} \\ 0 & A_{i,t+1}^{(1)}\cdots A_{i,2t-1}^{(1)} \end{bmatrix}. \tag{133}$$

在对 $i=1, 2, \cdots, t$ 作完上述分解后，经过适当的行交换，由于矩阵 $[A_{ij}]$ 中零块的特殊分布，Q^TA 呈下形：①

$$Q^TA=\begin{bmatrix} R_1 & & A_{1,t+1}^{(0)} & \cdots\cdots & A_{1,(3t/2)+1}^{(0)} & \cdots & A_{1,2t-1}^{(0)} \\ & \ddots & \vdots & & \vdots & & \vdots \\ & & R_t\ A_{t,t+1}^{(0)} & \cdots\cdots & A_{t,(3t/2)+1}^{(0)} & \cdots & A_{t,2t-1}^{(0)} \\ & & \boxed{\begin{matrix} A_{1,t+1}^{(1)} \\ A_{2,t+1}^{(1)} \end{matrix}} & & \begin{matrix} A_{1,(3t/2)+1}^{(1)} \\ A_{2,(3t/2)+1}^{(1)} \end{matrix} & \begin{matrix} \cdots \\ \cdots \end{matrix} & \begin{matrix} A_{1,2t-1}^{(1)} \\ A_{2,2t-1}^{(1)} \end{matrix} \\ & & & \ddots & \vdots & & \vdots \\ & & & \boxed{\begin{matrix} A_{t-1,3t/2}^{(1)} \\ A_{t,3t/2}^{(1)} \end{matrix}} & \begin{matrix} A_{t-1,(3t/2)+1}^{(1)} \\ A_{t,(3t/2)+1}^{(1)} \end{matrix} & \begin{matrix} \cdots \\ \cdots \end{matrix} & \begin{matrix} A_{t-1,2t-1}^{(1)} \\ A_{t,2t-1}^{(1)} \end{matrix} \end{bmatrix}. \tag{134}$$

然后，用前述方法对新生成的主对角块

$$\boxed{\begin{matrix} A_{s,t+s}^{(1)} \\ A_{s+1,t+s}^{(1)} \end{matrix}}\ s=1,\ 2,\ \cdots,\ t/2,$$

实行三角化，完成后再进行行交换，把新的正方上三角阵向上

---

① 我们假定矩阵 $[A_{ij}]$ 所具有的特殊结构可以保证所述形式的分解得以实现。

平移．以后仿此进行，最后就得到了矩阵 A 的QR分解 $A=Q\begin{bmatrix}R\\0\end{bmatrix}$ 其中

$$R=\begin{bmatrix} R_1 & & & A_{1,t+1}^{(0)} & \cdots & A_{1,2t-1}^{(0)} \\ & \ddots & & \vdots & & \\ & & R_t & A_{t,t+1}^{(0)} & \cdots & A_{t,2t-1}^{(0)} \\ & & & & & \vdots \\ & & & R_{t+1} & & \vdots \\ & & & & \ddots & A_{2t-2,2t-1}^{(0)} \\ & & & & & R_{2t-1} \end{bmatrix} \tag{135}$$

这时，原超定线性方程组就化为

$$R\boldsymbol{x}=\boldsymbol{c},$$

这里的右端向量 $\boldsymbol{c}$ 为 m 维向量 $Q^T\boldsymbol{b}$ 的前 n 个向量．应当指出，在矩阵 A 分解过程中作行交换时，右端的分量也应作相应的交换，对每一个 A_i 为选主元而需作列交换时，$\boldsymbol{x}$ 的分量也应作相应的交换．

把向量 $\boldsymbol{x}$ 和右端向量 $\boldsymbol{c}$ 都按 R 中子块来分段，并记之为：

$$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{x}_1\\ \vdots\\ \boldsymbol{x}_{2t-1}\end{bmatrix},\quad \boldsymbol{c}=\begin{bmatrix}\boldsymbol{c}_1\\ \vdots\\ \boldsymbol{c}_{2t-1}\end{bmatrix}, \tag{136}$$

其中，$\boldsymbol{x}_i$ 和 $\boldsymbol{c}_i$ 都是 n_i 维向量，$\sum n_i=n$．

利用回代，依次解三角方程组：

1° $R_{2t-1}\boldsymbol{x}_{2t-1}=\boldsymbol{c}_{2t-1}$；

2° $R_i\boldsymbol{x}_i=\boldsymbol{c}_i-\sum\limits_{j=i+1}^{2t-1}A_{ij}^{(0)}\boldsymbol{x}_j,\ i=2t-2,\ 2t-3,\cdots,t+1$；

3° $R_i\boldsymbol{x}_i=\boldsymbol{c}_i-\sum\limits_{j=t+1}^{2t-1}A_{ij}^{(0)}\boldsymbol{x}_j,\ i=t,\ t-1,\ \cdots\cdots,\ 1.$

最后就得到了所需要的解．

6-3 LU 分解法

本章开头所介绍的 LU 分解法，即，Gauss 消去法，也可以用来解这一类型的最小二乘问题．今简述其步骤于下．

第一步：对每一个主对角块 A_i 作 LU 分解，但选主元只限于在 A_i 内进行以免破坏矩阵的结构．设

$$P_1A_iP_2=L_iU_i,$$

其中 L_i 为下梯形阵，U_i 为上三角阵，它们的形式为：

$L_i=$ [下梯形阵示意图：1 … 1，0]，$U_i=$ [上三角阵示意图，0]．

P_1, P_2 为两个排列阵．

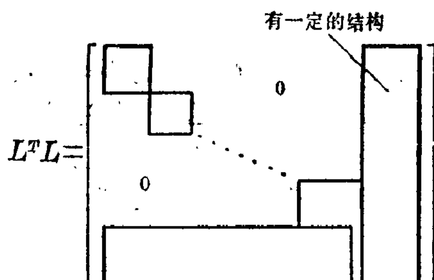
和前法一样，需要把分解过的子块移位和合并．最后就得到了矩阵 A 的 LU 分解

$$P^{(1)}AP^{(2)}=LU,$$

而 U 的形式为

$U=$ [示意图：有一定的结构；0；0]，

L 是一个分块上梯形阵，L^TL 的形式为



第二步：用分块 Cholesky 分解法解方程组

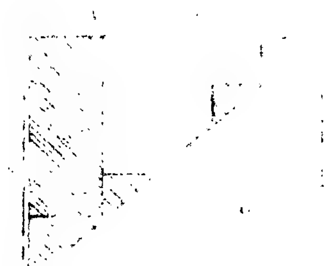
$$L^TLw=L^TP^{(1)}b$$

第三步：利用回代解方程组

$$Uz=w.$$

第四步：令 $x=P^{(2)T}z$，就得到了所需要的解.

最后我们指出，在前述两种算法中，QR 分解法具有较好的数值稳定性，如果矩阵 A 的条件数不太大，得到的结果相当精确，而且，在增、减变量个数和观测次数时，也便于处理. 关于 LU 分解法，从数值稳定性来说，不如 QR 分解，但是却优于通常直接解法方程的方法. 而且，一般说来，此法比直交化方法更有利于保持矩阵的稀疏性.



附录 1

方阵的 Drazin 逆

§1 引　　言

在本书第 2 章中，我们从线性算子方程求解问题出发，引进了一类广义逆算子，从而建立了一类相应的广义逆矩阵的基本理论．这些广义逆矩阵保留了非奇异方阵之逆的若干性质，但也失掉了另一些性质．例如，若 A 为 n 阶非奇异方阵，则它的逆矩阵 A^{-1} 具有下列性质：

(i) $AA^{-1}=A^{-1}A$,

(ii) $(A^{-1})^p=(A^p)^{-1}$,

(iii) $A^{p+1}A^{-1}=A^p$, 　p 为正整数

(iv) $\lambda\in\sigma(A)$ 必须且只需 $\lambda^+\in\sigma(A^{-1})$,

这里的 $\sigma(A)$ 表示矩阵 A 的特征值集合，并且

$$\lambda^+=\begin{cases}\lambda^{-1}, & 若\ \lambda\neq 0,\\ 0, & 若\ \lambda=0.\end{cases}$$

广义逆矩阵 A^-, A_l^-, A_m^- 和 A^+，一般不具有这些性质．在这个附录里我们将介绍方阵的另一种广义逆，从某种角度来看，它和通常的逆矩阵更为相似．这类广义逆就叫做矩阵的 Drazin 逆，或 Drazin 逆矩阵．这种广义逆矩阵在奇异线性常微分方程组以及奇异线性差分方程组求解问题上有重要应用．

和第 2 章一样，我们先考虑线性算子的 Drazin 逆，然后再考虑 Drazin 逆矩阵．在本附录里，我们所考虑的线性算

子，都是从空间 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 的线性算子，所以它们的矩阵表示都是 n 阶方阵.

§2 线性算子的结构

2-1 线性算子的指标及空间的相应不变子空间直接和分解

我们首先证明一个引理.

引理 1 对于任何线性算子 $\boldsymbol{A}$，总存在一个最小的非负整数 k，使

$$\begin{aligned}&\mathscr{R}(\boldsymbol{A}^\circ)\supset\mathscr{R}(\boldsymbol{A})\supset\cdots\supset\mathscr{R}(\boldsymbol{A}^k)=\mathscr{R}(\boldsymbol{A}^{k+1})=\cdots,\\&\mathscr{N}(\boldsymbol{A}^\circ)\subset\mathscr{N}(\boldsymbol{A})\subset\cdots\subset\mathscr{N}(\boldsymbol{A}^k)=\mathscr{N}(\boldsymbol{A}^{k+1})=\cdots,\end{aligned}\tag{1}$$

且

$$\mathscr{R}(\boldsymbol{A}^k)\cap\mathscr{N}(\boldsymbol{A}^k)=\{\boldsymbol{\theta}\}.\tag{2}$$

这里我们约定 $\boldsymbol{A}^\circ=\boldsymbol{I}$ 为单位算子.

证明 关系 (1) 等价于

$$\begin{aligned}&\dim\mathscr{R}(\boldsymbol{A}^\circ)>\dim\mathscr{R}(\boldsymbol{A})>\cdots>\dim\mathscr{R}(\boldsymbol{A}^k)\\&\quad=\dim\mathscr{R}(\boldsymbol{A}^{k+1})=\cdots\end{aligned}$$

$$\begin{aligned}&\dim\mathscr{N}(\boldsymbol{A}^\circ)<\dim\mathscr{N}(\boldsymbol{A})<\cdots<\dim\mathscr{N}(\boldsymbol{A}^k)\\&\quad=\dim\mathscr{N}(\boldsymbol{A}^{k+1})=\cdots.\end{aligned}$$

因为

$$0\leqslant\dim\mathscr{R}(\boldsymbol{A}^{i+1})\leqslant\dim(\boldsymbol{A}^i)\leqslant n,$$

$$n\geqslant\dim\mathscr{N}(\boldsymbol{A}^{i+1})\geqslant\dim\mathscr{N}(\boldsymbol{A}^i)\geqslant0,$$

对 $i=0,1,\cdots$，所以必存在最小的非负整数 k 使

$$\dim\mathscr{R}(\boldsymbol{A}^k)=\dim\mathscr{R}(\boldsymbol{A}^{k+1}).$$

于是，对任何 $i>1$，显然有

$$\mathscr{R}(\boldsymbol{A}^{k+i})=\mathscr{R}(\boldsymbol{A}^{k+1}),$$

从而有

$$\dim \mathscr{R}(\boldsymbol{A}^{k+i}) = \dim \mathscr{R}(\boldsymbol{A}^{k+1}) = \dim \mathscr{R}(\boldsymbol{A}^{k}).$$

由于

$$\dim \mathscr{R}(\boldsymbol{A}^{k}) + \dim \mathscr{N}(\boldsymbol{A}^{k}) = n,$$

据此可以证明第二个关系式也成立.

现在我们来证明 (2)；即子空间 $\mathscr{R}(\boldsymbol{A}^{k})$ 和 $\mathscr{N}(\boldsymbol{A}^{k})$ 中唯一的共同元素是零向量. 为此设

$$\boldsymbol{x} \in \mathscr{R}(\boldsymbol{A}^{k}) \cap \mathscr{N}(\boldsymbol{A}^{k}),$$

则必存在向量 $\boldsymbol{y} \in \mathscr{C}^{n}$ 使

$$\boldsymbol{x} = \boldsymbol{A}^{k}\boldsymbol{y}.$$

但 $$\boldsymbol{A}^{k}\boldsymbol{x} = \boldsymbol{A}^{2k}\boldsymbol{y} = \boldsymbol{\theta},$$

这就意味着 $$\boldsymbol{y} \in \mathscr{N}(\boldsymbol{A}^{2k}) = \mathscr{N}(\boldsymbol{A}^{k}).$$

所以必有 $$\boldsymbol{x} = \boldsymbol{\theta}.$$

于是引理得证.

定义 1 引理中的最小非负整数 k 称为线性算子 $\boldsymbol{A}$ 的指标 (Index), 记作

$$\mathrm{Ind}(\boldsymbol{A}) = k.$$

线性算子 $\boldsymbol{A}$ 所确定的空间 $\mathscr{C}^{n}$ 的直接和分解

因为

$$\boldsymbol{A}\mathscr{R}(\boldsymbol{A}^{k}) = \mathscr{R}(\boldsymbol{A}^{k+1}) = \mathscr{R}(\boldsymbol{A}^{k}),$$
$$\boldsymbol{A}\mathscr{N}(\boldsymbol{A}^{k}) = \mathscr{N}(\boldsymbol{A}^{k+1}) = \mathscr{N}(\boldsymbol{A}^{k}),$$

所以, $\mathscr{R}(\boldsymbol{A}^{k})$ 和 $\mathscr{N}(\boldsymbol{A}^{k})$ 都是线性算子 $\boldsymbol{A}$ 的不变子空间. 此外, 由于 $\mathscr{R}(\boldsymbol{A}^{k}) \cap \mathscr{N}(\boldsymbol{A}^{k}) = \{\boldsymbol{\theta}\}$, 从而有直接和分解

$$\mathscr{C}^{n} = \mathscr{R}(\boldsymbol{A}^{k}) \oplus \mathscr{N}(\boldsymbol{A}^{k}). \tag{3}$$

我们称 (3) 为空间 $\mathscr{C}^{n}$ 由线性算子 $\boldsymbol{A}$ 所确定的不变子空间直接和分解.

2-2 线性算子的分解

据空间 $\mathscr{C}^{n}$ 的直接和分解 (3) 可知, 任何一个向量 $\boldsymbol{x} \in \mathscr{C}^{n}$, 都可以唯一地表示成

$$x=u+v,$$

其中 $u\in\mathscr{R}(A^k)$, $v\in\mathscr{N}(A^k)$. 今规定算子 A_1, A_2 如下:

$$A_1x=Au,\quad A_2x=Av. \tag{4}$$

显然可知, A_1 和 A_2均为线性算子, 且

$$A_1v=\theta,\quad A_2u=\theta, \tag{5}$$

从而有,

$$(A_1+A_2)x=Au+Av=A(u+v)=Ax.$$

由于上式对任何 x 都成立, 所以有

$$A=A_1+A_2. \tag{6}$$

这就是说, 线性算子 A 可以表示成线性算子 A_1 与 A_2 之和.

我们称 (6) 为空间 $\mathscr{C}^n$ 的直接和分解 (3) 所确定的线性算子 A 的分解.

线性算子 A_1 和 A_2 具有一些重要性质.

定理1 对于线性算子 A 的分解 (6), 我们有

$$\operatorname{Ind}(A_1)\leqslant 1,\quad A_1A_2=A_2A_1=0, \tag{7}$$

并且, A_2 是一个 k 次幂零算子, 即

$$A_2^k=0,\quad A_2^{k-1}\neq 0, \tag{8}$$

其中 $k=\operatorname{Ind}(A)$.

证明 我们分两种情形来考虑.

1) $\dim\mathscr{R}(A)=n$.

这时, $\operatorname{Ind}(A)=0$, 即 $k=0$. 由于

$$\mathscr{R}(A^\circ)=\mathscr{C}^n,\quad \mathscr{N}(A^\circ)=\{\theta\},$$

所以 $\qquad A_1=A,\ A_2=0,\ A_1A_2=A_2A_1=0.$

2) $\dim\mathscr{R}(A)<n$.

这时 $k=\operatorname{Ind}(A)>0$. 容易证明 $\mathscr{R}(A_1)=\mathscr{R}(A_1^2)$, 所以, $\operatorname{Ind}(A_1)=1$, 对任意的

$$x=u+v,\quad u\in\mathscr{R}(A^k),\quad v\in\mathscr{N}(A^k),$$

$$A_2^i x = (A - A_1)^i x = (A - A_1)^{i-1}(A - A_1)(u + v)$$
$$= (A - A_1)^{i-1} A v = \cdots = A^i v.$$

由此可知 $A_2^k x = \theta$，所以

$$A_2^k = 0,$$

而当 $i < k$ 时，$A_2^i \neq 0$. 所以 A_2 是 k 次幂零算子，另外，

$$A_1 A_2 x = A_1 A_2 (u + v) = A_1 A v = \theta,$$

同理有
$$A_2 A_1 x = A_2 A u = \theta,$$

从而可知
$$A_1 A_2 = A_2 A_1 = 0.$$

于是定理得证.

定理 2 在前述定理中的条件下，分解式 (6) 是唯一的. 即，若 $A = B_1 + B_2$ 为另一种分解，且

$$B_1 B_2 = B_2 B_1 = 0,$$
$$\mathrm{Ind}(B_1) \leqslant 1,\ B_2 \text{ 为 } k \text{ 次幂零矩阵},$$

则必有
$$B_1 = A_1,\ B_2 = A_2.$$

证明从略.

定义 2 我们称分解式 (6) 中的 A_1 为线性算子 A 的核心 (core) 部分，称 A_2 为 A 的幂零 (nillpotent) 部分.

用归纳法可以证明，对任何正整数 p，恒有

$$A^p = A_1^p + A_2^p. \tag{9}$$

§3 线性算子的 Drazin 逆及其基本性质

3-1 Drazin 逆算子的定义

设
$$A = A_1 + A_2,$$

为线性算子 A 的核心—幂零分解. 据算子 A_1 的定义可知，在子空间 $\mathscr{R}(A^k)$ 上 A_1 和 A 恒等，并且

$$A_1\mathscr{R}(A^k)=\mathscr{R}(A^k),$$

这就是说，A_1 在 $\mathscr{R}(A^k)$ 与其自身之间建立了1—1对应关系．所以，在 $\mathscr{R}(A^k)$ 上，算子 A_1 存在有逆算子．记 A_I 为只定义在 $\mathscr{R}(A^k)$ 上并在其上与 A 恒等的线性算子．我们有

定义3 设 A 为给定的线性算子，今规定算子 A^D 如下：对任何 $x=u+v\in\mathscr{C}^n$，$u\in\mathscr{R}(A^k)$，$v\in\mathscr{N}(A^k)$，$k=\mathrm{Ind}(A)$，

$$A^Dx=A_I^{-1}u. \tag{10}$$

则 A^D 为线性算子，称之为 A 的 Drazin 逆算子．

据算子 A^D 的定义可知，若 $x\in\mathscr{N}(A^k)$，则必有

$$A^Dx=\theta,$$

并且，对任意的 $u\in\mathscr{R}(A^k)$，恒有

$$AA^Du=u,\quad A^DAu=u,$$

而对任意的 $v\in\mathscr{N}(A^k)$，恒有

$$AA^Dv=A^DAv=\theta.$$

这就表示，对一切 $x\in\mathscr{C}^n$，恒有

$$AA^Dx=A^DAx,$$

从而可知

$$AA^D=A^DA. \tag{11}$$

此外还不难看出，算子 AA^D 或 A^DA，实际上就是空间分解(3)所确定的沿子空间 $\mathscr{N}(A^k)$ 到子空间 $\mathscr{R}(A^k)$ 上的投影算子，即

$$AA^D=A^DA=P_{\mathscr{R}(A^k),\mathscr{N}(A^k)}. \tag{12}$$

进而可知

$$I-AA^D=I-A^DA=P_{\mathscr{N}(A^k),\mathscr{R}(A^k)}. \tag{13}$$

3-2 Drazin 逆算子的特征性质

根据线性算子 Drazin 逆的定义，我们已经证明 A^D 和 A 可交换，即关系(11)成立．现在我们来证明

定理 3 线性算子 $\boldsymbol{A}$ 的 Drazin 逆 A^D 满足下列关系:

$$\boldsymbol{A}^D\boldsymbol{A}\boldsymbol{A}^D=\boldsymbol{A}^D, \tag{14-1}$$

$$\boldsymbol{A}^{k+1}\boldsymbol{A}^D=\boldsymbol{A}^k, \tag{14-2}$$

其中 $k=\mathrm{Ind}(\boldsymbol{A})$.

证明 对任意的向量 $\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v}\in\mathscr{C}^n$, 其中 $\boldsymbol{u}\in\mathscr{R}(\boldsymbol{A}^k)$, $\boldsymbol{v}\in\mathscr{N}(\boldsymbol{A}^k)$, 恒有

$$\begin{aligned}\boldsymbol{A}^D\boldsymbol{A}\boldsymbol{A}^D\boldsymbol{x}&=\boldsymbol{A}^D\boldsymbol{A}\boldsymbol{A}_I^{-1}\boldsymbol{u}=\boldsymbol{A}^D\boldsymbol{u}\\&=\boldsymbol{A}^D(\boldsymbol{u}+\boldsymbol{v})=\boldsymbol{A}^D\boldsymbol{x},\end{aligned}$$

故有 $\boldsymbol{A}^D\boldsymbol{A}\boldsymbol{A}^D=\boldsymbol{A}^D$. 又

$$\begin{aligned}\boldsymbol{A}^{k+1}\boldsymbol{A}^D\boldsymbol{x}&=\boldsymbol{A}^k(\boldsymbol{A}\boldsymbol{A}^D)\boldsymbol{u}\\&=\boldsymbol{A}^k\boldsymbol{u}=\boldsymbol{A}^k(\boldsymbol{u}+\boldsymbol{v})=\boldsymbol{A}^k\boldsymbol{x},\end{aligned}$$

从而有 $\boldsymbol{A}^{k+1}\boldsymbol{A}^D=\boldsymbol{A}^k$. 证毕.

此外, 进一步还可以证明

定理 4 设 $\boldsymbol{A}$, $\boldsymbol{G}$ 均为从 $\mathscr{C}^n$ 到 $\mathscr{C}^n$ 的线性算子, $k=\mathrm{Ind}(\boldsymbol{A})$, 欲 $\boldsymbol{G}=\boldsymbol{A}^D$, 其充分必要条件为:

$$\begin{aligned}&(\text{i})\ \boldsymbol{G}\boldsymbol{A}\boldsymbol{G}=\boldsymbol{G},\\&(\text{ii})\ \boldsymbol{G}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{G},\\&(\text{iii})\ \boldsymbol{A}^{k+1}\boldsymbol{G}=\boldsymbol{A}^k.\end{aligned} \tag{15}$$

证明 条件的必要性已经证明, 现在只需要证明条件的充分性就可以了. 现在我们来证明, 满足条件 (15) 的线性算子 $\boldsymbol{G}$ 必具有下列性质:

对任意的 $\boldsymbol{x}=\boldsymbol{u}+\boldsymbol{v}$, 其中 $\boldsymbol{u}\in\mathscr{R}(\boldsymbol{A}^k)$, $\boldsymbol{v}\in\mathscr{N}(\boldsymbol{A}^k)$, 恒有

$$\begin{aligned}\boldsymbol{A}\boldsymbol{G}\boldsymbol{u}&=\boldsymbol{u},\\\boldsymbol{A}\boldsymbol{G}\boldsymbol{v}&=\boldsymbol{\theta}.\end{aligned} \tag{16}$$

因为

$$\boldsymbol{A}^{k+1}\boldsymbol{G}\boldsymbol{x}=\boldsymbol{A}^k\boldsymbol{x}=\boldsymbol{A}^k\boldsymbol{u}=\boldsymbol{A}^k\boldsymbol{A}\boldsymbol{G}\boldsymbol{u},$$

即
$$A^k(u-AGu)=\theta,$$
这就表示 $u-AGu\in\mathcal{N}(A^k)$. 而 $AGu=(AG)^k u=A^kG^ku$,
所以
$$u-AGu\in\mathcal{R}(A^k),$$
从而恒有
$$AGu=u,$$
即 (16) 之一成立. 另一方面
$$Gv=GAGv=\cdots=G(AG)^kv$$
$$=G^{k+1}A^kv=\theta,$$
即 (16) 之二也成立. 于是必有
$$G=A^D. \qquad 证毕.$$

(15) 中的 (i)～(iii) 便是 Drazin 逆的特征性质. 应当指出, 在 (iii) 中把 k 换成任一个 $p\geqslant k$, 对 Drazin 逆来说, 关系 (iii) 仍成立.

3-3 算子核心部分的表示

线性算子 A 的核心 A_1 具有特殊的性质, 现在我们用 A 的 Drazin 逆 A^D 把 A_1 表示出来. 我们有

定理 5 $A_1=AA^DA.$ (17)

证明 令
$$B_1=AA^DA,\ B_2=A-AA^DA,$$
我们只需证明线性算子 B_1 和 B_2 满足定理 2 中的条件就行了.

若 $\mathrm{Ind}(A)=k=0$, 这时 A 为非奇异线性算子, 这时 $A^D=A^{-1}$, 定理显然成立. 今设 $k\geqslant1$. 因为
$$\mathcal{R}(AA^DA)=\mathcal{R}(A^2A^D)=A^2\mathcal{R}(A^D)$$
$$=A^2\mathcal{R}(A^k)=\mathcal{R}(A^k)=\mathcal{R}((AA^DA)^2),$$
从而可知 $\mathrm{Ind}(AA^DA)=1$, 即 $\mathrm{Ind}(B_1)=1$. 另外,
$$B_1B_2=AA^DA(A-AA^DA)=A^3A^D-A^4(A^D)^2$$
$$=A^3AA^DA-A^4(A^D)^2=0.$$

同理可证 $B_2B_1=0$.

现在我们来证明 B_2 是 k 次幂零算子. 对任意的 $x=u+v$, $u\in\mathscr{R}(A^k)$, $v\in\mathscr{N}(A^k)$, 恒有

$$B_2^kx=(A-AA^DA)^kx=A^k(I-AA^D)^k(u+v)$$
$$=A^kv=\theta,$$

所以必有 $B_2^k=0$. 但,存在有 $v\in\mathscr{N}(A^k)$ 而 $v\notin\mathscr{N}(A^{k-1})$, 对这样的 v,

$$B_2^{k-1}v=(A-AA^DA)^{k-1}v=A^{k-1}(I-A^DA)^{k-1}v$$
$$=A^{k-1}v\neq\theta,$$

从而可知 $B_2^{k-1}\neq 0$. 于是我们便证明 B_2 是一个 k 次幂零算子. 因此, $B_1=A_1$, 即

$$A_1=AA^DA.$$

定理得证.

为了把线性算子 A 的核心-幂零分解表示得更为清楚, 我们把 A 的核心部分(即 A_1)记成 C_A, 而把 A 的幂零部分记为 N_A(即 A_2). 于是有

$$CA=AA^DA,\quad NA=A-CA. \tag{18}$$

不难证明, 对任意的正整数 p, 线性算子 A^p 的核心-幂零分解为:

$$A^p=C_A^p+N_A^p, \tag{19}$$

其中的 C_A^p 为 A^p 的核心,而 N_A^p 为 A^p 的幂零部分.

根据我们用来表示核心和幂零部分的记号, (19) 可以表示成:

$$C_A^p=CA_p,\quad N_A^p=NA_p. \tag{20}$$

此外,我们还不难证明,对任意正整数 p, 恒有

$$(A^p)^D=(A^D)^p. \tag{21}$$

因为, 一方面空间 $\mathscr{C}^n$ 由算子 A 和 A^p 所确定的不变子空间分解是相同的,另一方面,对任意的

$$x=u+v\in\mathscr{C}^n,\ u\in\mathscr{R}(A^k),\ v\in\mathscr{N}(A^k),$$
$$k=\mathrm{Ind}(A),$$

我们恒有

$$(A^D)^p x=(A^D)^{p-1}A^D x=(A^D)^{p-1}A_1^{-1}u$$
$$=\cdots\cdots=(A_1^{-1})^p u=(A_1^p)^{-1}u=(A^p)^D x.$$

3-4 Drazin 逆的特例——群逆 (Group Inverse)

一般说来，线性算子 A 的 Drazin 逆 A^D 不满足关系

$$AA^DA=A,\tag{22}$$

只有在一定的条件下 A^D 才具有这种性质．我们有

定理 6 关系 (22) 成立的充分必要条件为：

$$\mathrm{Ind}(A)\leqslant 1.\tag{23}$$

证明 若 $\mathrm{Ind}(A)=0$，则 A 为非奇异，这时 $A^D=A^{-1}$，(22) 显然成立．设 $\mathrm{Ind}(A)=1$，这时，对任意的 $x=u+v$，$u\in\mathscr{R}(A)$，$v\in\mathscr{N}(A)$，必有

$$A_2x=Av=\theta,$$
$$A_1x=Au=A(u+v)=Ax.$$

从而可知 $N_A=A_2=0,\ C_A=A_1=A=AA^DA.$

如果 $\mathrm{Ind}(A)=k\geqslant 1$，欲 (22) 成立，必须

$$\mathscr{R}(A)=\mathscr{R}(AA^DA)=A^2\mathscr{R}(A^D)=A^2\mathscr{R}(A^k)$$
$$=\mathscr{R}(A^{k+2}),$$

从而必有 $\mathscr{R}(A)=\mathscr{R}(A^2)$，即 $k=1$．证毕．

定义 4 若 $\mathrm{Ind}(A)\leqslant 1$，则称 A 的 Drazin 逆 A^D 为 A 的群逆，记作 $A^\#$.

根据群逆的定义可知，$A^\#$ 具有下列特征性质：

$$\begin{aligned}&(\mathrm{i})\ AA^\#A=A,\\&(\mathrm{ii})\ A^\#AA^\#=A^\#,\\&(\mathrm{iii})\ A^\#A=AA^\#.\end{aligned}\tag{24}$$

Drazin 逆 $\boldsymbol{A}^D$ 或群逆 $\boldsymbol{A}^\#$，一般并不是 $\boldsymbol{A}^+$，但是在一定条件下它们是 $\boldsymbol{A}^+$. 我们有

定理 7 如欲 $\boldsymbol{A}^D$ 或 $\boldsymbol{A}^\#$ 是 $\boldsymbol{A}^+$，其充分必要条件是：$\boldsymbol{A}$ 和 $\boldsymbol{A}^+$ 可交换，即

$$\boldsymbol{A}^+\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^+. \tag{25}$$

证明 若(25)成立，则 $\boldsymbol{A}^+$ 便具有 $\boldsymbol{A}^\#$ 的全部特征性质，所以这时必有

$$\boldsymbol{A}^+=\boldsymbol{A}^D=\boldsymbol{A}^\#.$$

反之，若 $\boldsymbol{A}^+=\boldsymbol{A}^D$，则必有

$$\boldsymbol{A}\boldsymbol{A}^D\boldsymbol{A}=\boldsymbol{A},$$

从而可知 $\boldsymbol{A}^+=\boldsymbol{A}^\#=\boldsymbol{A}^D$. 证毕.

§4 矩阵的 Drazin 逆

4-1 引言

设 $\boldsymbol{A}$ 是一个给定的从空间 $\mathscr{C}^n\to\mathscr{C}^n$ 的线性算子，在空间 $\mathscr{C}^n$ 中建立了坐标系以后，用第 1 章 **4-3** 中的方法，可以把它表示成一个唯一的 n 阶方阵 A. 反之，一个 n 阶方阵 A，也确定了一个相应的线性算子 $\boldsymbol{A}$，在给定的坐标系中，A 是 $\boldsymbol{A}$ 的矩阵表示. n 阶方阵和从 $\mathscr{C}^n\to\mathscr{C}^n$ 的线性算子之间是一一对应的. 不但如此，而且在所给定的坐标系中，线性算子 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的乘积 $\boldsymbol{AB}$，其矩阵表示为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 的矩阵表示 A 和 B 的乘积，线性组合 $\alpha\boldsymbol{A}+\beta\boldsymbol{B}$ 的矩阵表示为线性组合 $\alpha A+\beta B$. 其中，若 $\boldsymbol{A}$ 为非奇异，则逆算子 $\boldsymbol{A}^{-1}$ 的矩阵表示为 A^{-1}，单位算子 $\boldsymbol{I}$ 的矩阵表示为单位阵 I. 从线性算子和 n 阶方阵之间的这种对应关系出发，我们很容易建立 Drazin 逆矩阵的概念和有关性质.

定义 5 设 n 阶方阵 A 是线性算子 $\boldsymbol{A}$ 在给定坐标系中的矩阵表示，记 Drazin 逆算子 $\boldsymbol{A}^D$ 在同一坐标系中的矩阵表示为 A^D，我们称 A^D 为矩阵 A 的 Drazin 逆.

因为线性算子 $\boldsymbol{A}$ 的 Drazin 逆 $\boldsymbol{A}^D$ 是唯一的，它在给定坐标系中的矩阵表示也是唯一的，所以，矩阵 A 的 Drazin 逆矩阵 A^D 也是唯一的.

根据 Drazin 逆矩阵和 Drazin 逆算子之间的上述对应关系，为建立 Drazin 逆算子所引进的一些概念以及关于 Drazin 逆算子的性质，都不难照搬到 Drazin 逆矩阵上来. 应当指出，对给定的线性算子 $\boldsymbol{A}$，在空间 $\mathscr{C}^n$ 中建立了一种特殊的坐标系后，线性算子 $\boldsymbol{A}$ 及其 Drazin 逆 $\boldsymbol{A}^D$ 的矩阵表示，都具有非常简单的结构，从而使 Drazin 逆的性质变得十分明显. 同时，在实际应用中，常常不是直接应用抽象的线性算子，而是应用表示线性算子的矩阵，从实际应用的角度来考虑，有必要对矩阵的 Drazin 逆作进一步的研究.

4-2 线性算子在其不变子空间直接和上的矩阵表示

首先我们在空间 $\mathscr{C}^n$ 中取两组坐标基底

$$\{\alpha_i\} \text{ 和 } \{\beta_j\},$$

并设线性算子 $\boldsymbol{A}$ 在这两个坐标系中的矩阵表示分别为 A 和 B. 次设这两个基底之间有下列关系:

$$\beta_j=\sum_{i=1}^{n} q_{ij}\alpha_i,\quad \alpha_i=\sum_{j=1}^{n} p_{ji}\beta_j. \tag{26}$$

记

$$P=[p_{ij}],\ Q=[q_{ij}]. \tag{27}$$

不难证明矩阵 A 和 B 之间有下述关系:

$$A=QBP, \tag{28}$$

并且 P 和 Q 均为非奇异，且

$$Q=P^{-1} \text{ 或 } P=Q^{-1}. \tag{29}$$

从而可以把关系(28)表示成

$$A=P^{-1}BP \tag{30}$$

或

$$B=Q^{-1}AQ. \tag{31}$$

不难看出，矩阵 P 中的第 i 个列向量实际上就是向量 α_i 在坐标系 $\{\beta_j\}$ 中的坐标向量，而 Q 中的第 j 个列向量则是 β_j 在坐标系 $\{\alpha_i\}$ 中的坐标向量.

上述讨论说明，同一个线性算子在两个不同坐标系中的矩阵表示互为相似. 反之可知，两个相似矩阵可以看成是同一线性算子在不同坐标系中的表示.

现在我们假定空间 $\mathscr{C}^n$ 有直接和分解:

$$\mathscr{C}^n=\mathscr{W}_1\oplus\mathscr{W}_2, \tag{32}$$

其中的 $\mathscr{W}_1$ 和 $\mathscr{W}_2$ 都是线性算子 $\boldsymbol{A}$ 的不变子空间，即，我们有

$$\boldsymbol{A}\mathscr{W}_1\subset\mathscr{W}_1,\ \boldsymbol{A}\mathscr{W}_2\subset\mathscr{W}_2. \tag{33}$$

设

$$\dim\mathscr{W}_1=r, \tag{34}$$

则

$$\dim\mathscr{W}_2=n-r. \tag{35}$$

任一向量 $\boldsymbol{x}\in\mathscr{C}^n$ 都可以唯一地表示成

$$\boldsymbol{x}=\boldsymbol{w}_1+\boldsymbol{w}_2,\ \boldsymbol{w}_1\in\mathscr{W}_1,\ \boldsymbol{w}_2\in\mathscr{W}_2. \tag{36}$$

今定义线性算子 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_2$ 如下:

$$\boldsymbol{A}_1\boldsymbol{x}=\boldsymbol{A}\boldsymbol{w}_1,\ \boldsymbol{A}_2\boldsymbol{x}=\boldsymbol{A}\boldsymbol{w}_2. \tag{37}$$

于是显然有

$$\begin{aligned}\boldsymbol{A}_1\boldsymbol{w}_1=\boldsymbol{A}\boldsymbol{w}_1\in\mathscr{W}_1,\ \boldsymbol{A}_2\boldsymbol{w}_2=\boldsymbol{A}\boldsymbol{w}_2\in\mathscr{W}_2,\\ \boldsymbol{A}_1\boldsymbol{w}_2=\boldsymbol{A}_2\boldsymbol{w}_1=\boldsymbol{\theta}.\end{aligned} \tag{38}$$

从而，线性算子 $\boldsymbol{A}$ 可以表示成线性算子 $\boldsymbol{A}_1$ 与 $\boldsymbol{A}_2$ 之和:

$$\boldsymbol{A}=\boldsymbol{A}_1+\boldsymbol{A}_2. \tag{39}$$

令 $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{x}$，则向量 $\boldsymbol{y}$ 可以分解为向量 $\boldsymbol{y}_1$ 与 $\boldsymbol{y}_2$ 之和

$$\boldsymbol{y}=\boldsymbol{y}_1+\boldsymbol{y}_2, \tag{40}$$

其中

$$\boldsymbol{y}_1=\boldsymbol{A}_1\boldsymbol{x}=\boldsymbol{A}\boldsymbol{w}_1\in\mathscr{W}_1,\ \boldsymbol{y}_2=\boldsymbol{A}_2\boldsymbol{x}=\boldsymbol{A}\boldsymbol{w}_2\in\mathscr{W}_2. \tag{41}$$

在子空间 $\mathscr{W}_1$ 和 $\mathscr{W}_2$ 中分别取基底

$$\{\boldsymbol{p}_1,\ \cdots,\ \boldsymbol{p}_r\}\ 和\ \{\boldsymbol{p}_{r+1},\ \cdots,\ \boldsymbol{p}_n\}, \tag{42}$$

于是，

$$\{\boldsymbol{p}_1,\ \cdots,\ \boldsymbol{p}_n\} \tag{43}$$

便构成空间 $\mathscr{C}^n$ 的一个基底，我们引进下列坐标向量：

$\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 为表示向量 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 在基底(43)下的坐标向量；

$\boldsymbol{\xi}^{(1)}$ 和 $\boldsymbol{\xi}^{(2)}$ 为表示向量 $\boldsymbol{w}_1$ 和 $\boldsymbol{w}_2$ 在(42)中相应基底下的坐标向量；

η_1 和 η_2 为表示向量 $\boldsymbol{y}_1$ 和 $\boldsymbol{y}_2$ 在(42)中相应基底下的坐标向量；

于是，$\boldsymbol{\xi}$ 和 $\boldsymbol{\eta}$ 为 n 维，$\boldsymbol{\xi}^{(1)}$，$\boldsymbol{\eta}^{(1)}$ 为 r 维，$\boldsymbol{\xi}^{(2)}$，$\boldsymbol{\eta}^{(2)}$ 为 $n-r$ 维，并且有下列关系：

$$\begin{aligned}\boldsymbol{\eta}&=\begin{bmatrix}\boldsymbol{\eta}^{(1)}\\ \boldsymbol{\theta}_{n-r}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\theta}_r\\ \boldsymbol{\eta}^{(2)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\eta}^{(1)}\\ \boldsymbol{\eta}^{(2)}\end{bmatrix},\\ \boldsymbol{\xi}&=\begin{bmatrix}\boldsymbol{\xi}^{(1)}\\ \boldsymbol{\theta}_{n-r}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\theta}_r\\ \boldsymbol{\xi}^{(2)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\xi}^{(1)}\\ \boldsymbol{\xi}^{(2)}\end{bmatrix},\end{aligned} \tag{44}$$

其中的 $\boldsymbol{\theta}_r$ 和 $\boldsymbol{\theta}_{n-r}$ 分别为 r 维和 $n-r$ 维零向量。

记线性算子 $\boldsymbol{A}$ 在(42)中相应基底下在子空间 $\mathscr{W}_1$ 和 $\mathscr{W}_2$ 上的矩阵表示分别为 r 阶方阵 W_1 和 $n-r$ 阶方阵 W_2。记线性算子 $\boldsymbol{A}$ 在基底(43)下在空间 $\mathscr{C}^n$ 上的矩阵表示为 n 阶方阵 A，则有

$$\begin{aligned}\boldsymbol{\eta}=A\boldsymbol{\xi}&=\begin{bmatrix}W_1\boldsymbol{\xi}^{(1)}\\ W_2\boldsymbol{\xi}^{(2)}\end{bmatrix}=\begin{bmatrix}\boldsymbol{\eta}^{(1)}\\ \boldsymbol{\eta}^{(2)}\end{bmatrix}\\ &=\begin{bmatrix}W_1 & 0\\ 0 & W_2\end{bmatrix}\begin{bmatrix}\boldsymbol{\xi}^{(1)}\\ \boldsymbol{\xi}^{(2)}\end{bmatrix}.\end{aligned} \tag{45}$$

由此可知

$$A=\begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix}. \tag{46}$$

如果这同一线性算子 $\boldsymbol{A}$ 在另一基底下的矩阵表示为 B，则必存在非奇异 n 阶方阵 P 使

$$B=PAP^{-1}=P\begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix}P^{-1}. \tag{47}$$

4-3 矩阵的核心-幂零分解

作为前一小节中空间直接和分解的特例，令

$$\mathscr{W}_1=\mathscr{R}(\boldsymbol{A}^k),\ \mathscr{W}_2=\mathscr{N}(\boldsymbol{A}^k), \tag{48}$$

在这里我们假定 $k=\mathrm{Ind}(\boldsymbol{A})$，$r=\dim\mathscr{R}(\boldsymbol{A}^k)$。为了表示这种情形的特殊性，我们把 (46) 中的 r 阶方阵 W_1 和 $n-r$ 阶方阵分别改记为 C 和 N。于是 (46) 就变成：

$$A=\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}. \tag{49}$$

公式 (49) 中的矩阵 A, C 和 N 具有下列特性：

1° 因为 $\boldsymbol{A}\mathscr{R}(\boldsymbol{A}^k)=\mathscr{R}(\boldsymbol{A}^k)$，所以矩阵 C 为非奇异 r 阶方阵；

2° 矩阵 N 满足关系

$$N^k=0,$$

而 $N^{k-1}\neq 0$。

3° $\mathrm{rank}(A^k)=\mathrm{rank}(A^{k+1})=r$，若 $0\leqslant p<k$，则

$$\mathrm{rank}(A^p)>r.$$

定义 6 设 A 是一个给定的方阵，我们称满足关系

$$\mathrm{rank}(A^k)=\mathrm{rank}(A^{k+1})$$

的最小非负整数 k 为矩阵 A 的指标，并记作

$$k=\mathrm{Ind}(A).$$

若存在正整数 p，使

$$A^{p-1}\neq 0,\ A^p=0,$$

则称 A 为 p 次幂零矩阵.

显然，若 A 非奇异，则 $\mathrm{Ind}(A)=0$，若 A 为奇异，则

$$\mathrm{Ind}(A)\geqslant 1.$$

此外，不难证明，线性算子 $\boldsymbol{A}$ 及其在任何基底下的矩阵表示 A，有共同的指标，即恒有

$$\mathrm{Ind}(\boldsymbol{A})=\mathrm{Ind}(A).$$

并且，若 $\boldsymbol{A}$ 为 p 次幂零算子，则 A 也必为 p 次幂零矩阵.

根据前述讨论，我们有

定理 8 设 A 为任一 n 阶方阵，$\mathrm{Ind}(A)=k$，$\mathrm{rank}(A^k)=r$，则必存在非奇异的 n 阶方阵 P 和 r 阶方阵 C 以及 $n-r$ 阶 k 次幂零阵 N 使

$$A=P\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}P^{-1}. \tag{50}$$

附注 若 $k=0$，则 C 为 n 阶，这时 $N=0$.

公式(50)可改写成

$$A=P\begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix}P^{-1}+P\begin{bmatrix} 0 & 0 \\ 0 & N \end{bmatrix}P^{-1}, \tag{51}$$

令

$$C_A=P\begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix}P^{-1},\ N_A=P\begin{bmatrix} 0 & 0 \\ 0 & N \end{bmatrix}P^{-1}.$$

定义 7 我们称 C_A 为矩阵 A 的核心部分，称 N_A 为矩阵 A 的幂零部分，而称 (51) 为矩阵 A 的核心-幂零分解.

矩阵的核心-幂零分解是唯一的.

4-4 矩阵的 Drazin 逆

对给定的 n 阶方阵 A，它有唯一的核心-幂零分解(51). 记

$$A^D=P\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}. \tag{52}$$

定义 8 我们称 (52) 所规定的矩阵 A^D 为矩阵 A 的 Drazin 逆.

n 阶方阵 A, 可以把它看成是某一线性算子 $\boldsymbol{A}$ 在空间 $\mathscr{C}^n$ 的标准基底 $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ 下的矩阵表示, 而 (52) 所定义的矩阵 A^D 实际上就是线性算子 $\boldsymbol{A}$ 的 Drazin 逆 $\boldsymbol{A}^D$ 在同一空间基底下的矩阵表示. 在空间基底给定以后, n 阶方阵和从 $\mathscr{C}^n\to\mathscr{C}^n$ 的线性算子为一一对应. 这就是说, 矩阵的 Drazin 逆, 根据线性算子乘积的矩阵表示等于各乘子矩阵表示的积, 以及在给定坐标基底后线性算子矩阵表示的唯一性, §3 中所阐明的 Drazin 逆算子的性质, 都可以直接照搬到 Drazin 逆矩阵上来. 在本节中所引进的矩阵 A 的特殊分解 (51), 使相应的 Drazin 逆矩阵具有特殊的结构形式. 这对验证 Drazin 逆矩阵的性质带来了很多方便. 例如,

$$A^DAA^D=P\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}$$

$$=P\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}P^{-1}=A^D,$$

$$AA^DA=P\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}\begin{bmatrix} C^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} C & 0 \\ 0 & N \end{bmatrix}P^{-1}$$

$$=P\begin{bmatrix} C & 0 \\ 0 & 0 \end{bmatrix}P^{-1}=C_A.$$

如此等等. 今再列举一些关于 Drazin 逆矩阵的性质于下, 请读者试证之:

1° $\operatorname{Ind}(A^D)=\operatorname{Ind}(C_A)=\begin{cases}1, & \text{若 } \operatorname{Ind}(A)\geqslant 1, \\ 0, & \text{若 } \operatorname{Ind}(A)=0.\end{cases}$

2° $N_A C_A = C_A N_A = 0$.

3° $N_A A^D = A^D N_A = 0$.

4° $C_A A A^D = A A^D C_A = C_A$.

5° $A = C_A$，必须且只需 $\mathrm{Ind}(A) \leqslant 1$.

6° $(A^D)^D = C_A$.

7° $((A^D)^D)^D = A^D$.

8° $A^D = C_A^D$.

9° $(A^D)^* = (A^*)^D$.

10° $\lambda \in \sigma(A)$，必须且只需 $\lambda^+ \in \sigma(A^D)$.

附 录 2

关于广义逆乘积公式证明的补充

为了证明第 2 章 **5-4** 中的条件(iii)～(iv)成立时，广义逆的乘积公式成立，我们先证明

引理 设 A 和 G 分别为 $m\times n$ 和 $n\times m$ 矩阵。若

$$GA=A^{+}A,\ AG=AA^{+},$$

且

$$\operatorname{rank}(G)\leqslant\operatorname{rank}(A),$$

则必有

$$G=A^{+}.$$

证明 因为 $GA=A^{+}A$，故知 $AGA=A$。所以为了证明 $G=A^{+}$，仅需证明 $GAG=G$ 就行了.

因为 $\operatorname{rank}(G)\leqslant\operatorname{rank}(A)=\operatorname{rank}(AGA)\leqslant\operatorname{rank}(GA)$，而 $\operatorname{rank}(G)\geqslant\operatorname{rank}(GA)$，故必有

$$\operatorname{rank}(G)=\operatorname{rank}(GA).$$

从而便知

$$\mathscr{R}(G)=\mathscr{R}(GA).$$

由此可知，必存在 $n\times m$ 阶矩阵 C，使

$$G=GAC.$$

两端分别乘以 GA，便得

$$GAG=GAGAC=GAC=G.$$

关于乘积公式的证明

已知

$$B^{+}A^{+}=B^{+}BB^{+}A^{+}AA^{+}=A^{+}B^{+*}(B^{*}A^{*})A^{*+}A^{+},$$

所以 $\operatorname{rank}(B^{+}A^{+})\leqslant\operatorname{rank}(B^{*}A^{*})=\operatorname{rank}(AB)$. 从而据(iii)和(iv)便证明了乘积公式

$$(AB)^{+}=B^{+}A^{+}.$$

参　考　文　献

第 1 章

[1] R. Courant and D. Hilbert, Methods of Mathematical Physics, Vol.1.,1953(有中译本).

[2] B. Friedman, Principles and Techniques of Applied Mathematics, 1956.

[3] Л. А. Люстерник и В. И. Соболев, 泛函分析概要, 1951 (有中译本).

[4] 何旭初,线性代数,科学出版社,1978.

[5] P. R. Halmos, Finite-Dimensional Vector Spaces, 2-nd Edition, 1958, Van Norstrand-Reinhold, Princeton, New Jersey.

[6] Ф. Р. Гантмахер, 矩阵论, 1953(有中译本).

[7] G. Strang, Linear Algebra and its Applications, Academic Press, INC., USA. 1976.

第 2 章

[1] Ben-Isreal A., Greville T. N. E, Generalized Inverse Theory and Applications, Wiley, New York, 1974.

[2] Cambell S. L., Meyer, Jr C. D., Generalized Inverses of Linear Transformations, Pitman, 1979.

[3] Greville, T. N. E.,Note on the Generalized Inverse of a Matrix Product, SIAM Review, 8 (4), 1966.

[4] F. Kuhnert, Pseudoinverse Matrizen und die Methode der Regularisierung, Teubner Verlagsgeesellschaft, Leipzig, 1976.

[5] Moore E. H., On the Reciprocal of the General Algebraic Matrix (Abstract), Bull. Amer. Math. Soc., 26, 394—395, 1920.

[6] Penrose R., A Generalized Inverse for Matrices, Proc. Cam. Phil. Soc., 51 (1955), 406—413.

[7] Rao R., Matra S. K., Generalized Inverse of Matrices and its Applications, Wily, New York, 1971.

[8] M. Zuhair Nashed, Proceedings of an Advanced Seminar, Generalized Inverses and Applications, Academic Press, USA, 1976.

第 3 章

[1] A. Ben-Israel, On Error Bounds for the Generalized Inverse, SIAM J. Numer. Anal., 3 (1966) 585—592.

[2] T. N. E. Greville, Note On the Generalized Inverse of a Matrix Product, SIAM Review, 8 (4) 1966, 518—521.

[3] 何旭初,广义逆矩阵的连续性问题,高等学校计算数学学报 1 (2), 1979.

[4] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Printice-Hall 1974.

[5] B. Noble, Methods for Computing the Moore-Penrose Generalized Inverse, In «Generalized Inverse and Applications», Edited by M. Zuhair Nashed, 245—302, (1976).

[6] G. W. Stewart, On the Continuity of the Generalized Inverse, SIAM J. Appl. Math., 17 (1) 1969, 33—45.

[7] J. Stoer, Einführung in die Numerische Mathematik, I, Springer-Verlag Berlin Heidelberg, 1976.

[8] J. H. Wilkinson, The Algebraic Eigenvalue Problem, Clarendon Press, 1965.

第 4 章

[1] P. G. Ciarlet, Introduction à L'Analyse Numérique Matricielle et à L'Optimisation, MASSON, 1982.

[2] P. Deuflhard and W. Sautter, On Rank-Deficient Pseudoinverse, Linear Algebra and its Applications, 29, 91—111, 1980.

[3] Ф. Р. Гантмахер, 矩阵论, 1953.

[4] 何旭初,数值相关性理论及其应用,高等学校计算数学学报, 1 (1), 1979.

[5] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems. Printice Hall, 1974.

[6] J. Stoer Einführung in die Numerische Mathematik, I, Springer-Verlag Berlin Heidelberg, 1976.

[7] P. G. Ciarlet, Introduction à L' Analyse Numerique Matricielle et à L'Optimisation, MASSON, 1982.

[8] J. H. Wilkinson, The Algebraic Eigenvalue Problems, Clarendon Press, 1965.

[9] 同前：代数过程中的舍入误差(有中译本)。

第 5 章

[1] P. Deuflhard and W. Sautter, On Rank-Deficient Pseudoinverse, Linear Algebra and its Applications, 29, 91—111, 1980.

[2] I. S. Duff and J. K. Reid, A Comparison of Some Methods for the Solution of Sparse and Overdetermined Systems of Linear Equations, J. Inst. Math. Applics. 17, 267—280, 1976.

[3] P. E. Gill, G. H. Golub, W. Murray and M. A. Saunders, Methods for updating matrix factorization, Math, Comp, 28, 505—535, 1974.

[4] G. H. Golub, and R.J.Plemons (Knoville TN), Large-Scale Geodetic Least-Squares Adjustment by Dissection and Orthogonal Decomposition. Papers Presented at the Fourth International Symposium Linear Algebra and its Applications 34, 3—27(1980).

[5] G. H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, Numer. Math. 14, 403—420, 1970.

[6] R. J. Hanson and C. L. Lawson, Extensions and applications of the Householder algorithm for solving Linear least Squares problems, Math. Comp., 23, 787—812, 1969.

[7] H. Y. Huang, A direct method for the general solution of a system of linear equations, JOTA, 16(6), 429—445, 1975.

[8] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, Printice-Hall, 1974.

[9] G. Peters and J. H. Wilkinson, The least squares problom and pseudoinverse, The Computer J., 18 (3), 309—316, 1970.

[10] W. Sautter, Fehleranalyse für die Gauss-Elimination Zur Berechnung der Lösong minimaler Länge, Numer. Math., 30, 165—184, 1978.

[11] J. H. Wilkinson, The Algebraic Eigenvalue Problems, Clarendon Press, 1965.

[12] 赵金熙，相容线性方程组的 Huang 方法及其推广，高等学校计算数学学报，3 (1), 8—17, 1981.

[13] 张元继、林成森、何旭初，线性代数计算方法，科学出版社，1979.

[14] Ake Björck, Solving Linear Least Squares Problems by Gram-Schmidt Orthogonalization, BIT, 7, 1—21, 1967.

[15] M. L. Baart, The Use of Auto-corelation for Pseudorank Determina-

tion in Noisy Ill-Conditioned Linoar Leastsquaros Problems, SIAM J. Numer. Anal. 1982, 2, 241--247.

附 录 1

Campbell, S. L., Meyer Jr. C. D., Generalized Inverse of Linear Transformations. 1979, Pitman Publishing Limited.

索　引

B　部

Banach 引理　109
伴随算子　27
　~的性质　29
　~的矩阵表示　28
标准直交系　14
病态　109
　~程度的度量　116
补投影算子　34
保秩变形　101

C　部

残[illegible]　107
超平面　9
乘积的广义逆　67
Courant-Fischer 定理　84

D　部

等幂算子　33
等价矩阵　24
递推方法　168
　按列~　169
　按行~　171
单位坐标向量　7
大型稀疏最小二乘问题　175

F　部

范数　12
　向量的~　12
　矩阵的~　86
　线性算子的~　43
法方程　60
Frobenius 矩阵　148
非零解　5
非奇异方阵的摄动定理　91

G　部

Gauss 消去法　145
Gram 行列式　124
　~的性质　125
Gram-Schmidt 直交化方法　13
　改进的~　164
Greville 递推方法　163
广义逆算子　49
　~的定义　51, 53, 56, 62
　~的特征性质　53, 55, 61
　~的矩阵表示　70, 74, 75, 76
广义逆矩阵　72, 74, 75, 76

H　部

Hermite 矩阵　16
Hilbert 矩阵　111
Householder 变换　152
　Householder 矩阵　152

J　部

计算问题的病态和良态　109

极小范数解　47
极小最小二乘解　48
矩阵的保秩变形　101
矩阵的范数　86
矩阵的谱范数　88
矩阵的满秩分解　144
矩阵的奇异值　85
矩阵的奇异度　138

K　部

空间　2
　～的分解　10
　～的直交分解　17
　～的基底　4
　～的维数　5
　～的同构　7
　～的表示　6
亏秩矩阵广义逆的连续性问题　95

L　部

零元素　1
零向量　1

M　部

满秩分解　144
满秩矩阵广义逆的连续性　92, 94
矛盾方程　47
Moore 条件　64

N　部

内积　12
　～的表示　15

O　部

欧氏空间　13

P　部

谱范数　88
谱分解　39, 81
谱条件数　119
Penrose 条件　64

Q　部

QR 分解　154
QU 分解　154
QR-LL^T 分解　157
奇异值分解　86
奇异值扰动定理　88
奇异性　122
齐次线性方程组　5

S　部

Schwarz 不等式　42
算子乘积的广义逆　67
算法的数值稳定性　112
数值相关性　123
双向单值映射　7
Stewart 定理　100

T　部

同构　7
条件数　117
特征值的 min-max 和 max-min 性质　81

投影算子　33
　～的性质　32
　～的秩　39
　～的矩阵表示　36
　～的范数　45
　～的谱分解　170

W　部

伪秩　141

X　部

线性流形　49
线性向量空间　2
　～的定义　2
　～的基底　4
　～的维数　5
线性相关性　3
线性独立性　3
线性算子　18
　～的定义域　18
　～的零空间　20
　～的像空间　18
　～的运算　19
　～的矩阵表示　22
　～的范数　43
　～的秩　20
线性变换　24
线性算子方程　47
　～的相容性　47
向量　1
　～的内积　12
　～的范数　12, 41
　～的直交性　13
　～的表示　6
向后误差分析　115
相容性　47
相对相关性　131
　～指标　132
　～指标的计算方法　133
　～和条件数　140
选主元　146, 156

Y　部

酉空间　12
酉算子　30
　～的矩阵表示　32
拟条件数　139

Z　部

坐标　6
坐标向量　7
坐标变换公式　24
子空间　8
最小二乘解　48
最小二乘问题　56
　～的摄动定理　105
总体相关性　129
总体选主元　146
直交性　13
直交投影算子　34
　～的单秩分解(谱分解)　39
直交化方法　164
　～的数值稳定性　167
直交投影矩阵　76
直接和分解　10
自伴算子　30
　～的矩阵表示　30

[General Information]
□□=□□□□□□□□□□□□□□□
□□=□□□
□□=206
SS□=10653929
DX□=
□□□□=1985□11□□1□
□□□=□□□□□□□□□

1
§ 1
§ 2
§ 3
§ 4
§ 5
§ 6
§ 7
2
§ 1
§ 2 A-
§ 3 A
§ 4 A
§ 5 A+
§ 6
§ 7 Moore-Penrose
3
§ 1
§ 2 Banach
§ 3
§ 4
§ 5
§ 6

□4□ □□□□□□□□□□□□

§1 □□□□□□□□□□□□□

§2 □□□□□□□

§3 □□□□□□□□□□

§4 □□□□□□□□□□□□□

□5□ □□□□□□□□□□□□□□□□□□□

§1 □□□□□□□□□□□□

§2 Gauss□□□

§3 QR□QU□□□

§4 □□□□□

§5 □□□□□□□□□□□□□□□□□□□□

§6 □□□□□□□□□□□□□□□□□

□□1□□□Drazin□

□□2□□□□□□□□□□□□□□

□□□□

□□

□□□